UNIVERSITY OF TWENTE.

UNIVERSITAS GADJAH MADA

Dedicating Capacity and Scheduling Urgent Surgeries

Oki Almas Amalia

M.Sc. Thesis

VUmc

Supervisors:
Prof. Dr. Richard Boucherie
Dr. Ir. Maartje van de Vrugt
Ir. Jasper Bos
Dr. Irwan Endrayanto, MSc.

Preface

This report is the result of my graduation project which is a part of the dual degree master program in Applied Mathematics, University of Twente and Mathematics, Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia.

First, I would like to thank the chair of Stochastic Operations Research (SOR) group in University of Twente, Prof. Richard Boucherie and the master in mathematics program director of UGM, Prof. Supama along with the whole chair members of both parties who made the dual degree program possible. Many thanks to my daily supervisor, Maartje van de Vrugt, for giving me advice, new ideas, feedback and answering my (many) queries during my final project. Thank you to ir. Jasper Bos for providing the initial idea of the research and providing information on Jeroen Bosch Ziekenhuis. Also, thanks to Dr. Irwan Endrayanto for helping me settle everything in UGM. Thanks to dr. Rukmono Siswishanto, M. Kes., Sp.OG(K) for the help to make the cooperation with RSUP dr.Sardjito (Hospital), Yogyakarta in this research possible.

Thanks to all my committee members for taking out their time to read my thesis and for being a part of my graduation.

Thanks to all the group members at SOR and Center for Healthcare Operations Improvement and Research (CHOIR) for all the conversations at the lunch table (and walks) and the opportunity to join the biweekly meeting. It is amazing to have such a group of passionate people.

Finally, I would like to thank my family for supporting me during my study. Many thanks to my friends who inspired and encouraged me throughout the process. Thank you for being one call and one text away despite of the distance between us.

Yogyakarta, November 2019
Oki Almas Amalia

Abstract

The stochasticity of urgent patient arrivals provides a challenge in scheduling them into dedicated operating rooms (ORs). In our study, there are three categories of urgent patients: 30-minute, 6-hour, and 24-hour, distinguished by the maximum waiting time. We divide each day into 4 shifts, and we assume that the 30-minute patients are scheduled in the current shift. We determine the required capacity for the 6-hour and 24-hour patients using two queueing models: M/M/1/K and M/M/1 with priority. We perform a case study for various dedicated capacity levels and arrival rates. To schedule the admitted urgent patients in the dedicated capacity three Markov decision processes (MDP) based models are proposed. In the first MDP model we keep track of the target time of each admitted patient. For this model, it is optimal to treat the patients from the higher urgency levels first. In the next model, we modify the first model by allowing the OR manager to defer some 24-hour patients to another resource, e.g., elective ORs. While treating patients using the same policy as the previous model, the number of deferred patients depends on the costs. In the last model we assign newly arrived patients to a time-slot directly upon arrival. The optimal policy is to schedule 6-hour patients on the next shift, while 24-hour patients are scheduled somewhere before their deadline. As we cannot alter the assignment of the already scheduled patients to admit patients of higher urgency level, this model is less flexible than the other two models. The optimal policy of the proposed models boils down to simple rules that can be implemented easily by hospitals to treat urgent patients before their deadline.

Contents

Appendices

List of Tables

List of Figures

Chapter 1

Introduction

In a hospital, we can distinguish several types of surgeries based on the urgency level. Each urgency type has a maximum internal access time to the operating room (OR). The internal access time to the OR is defined as the time elapsed between the moment a doctor decides that a patient should have surgery and the moment the patient goes to the OR for the surgery [2]. In this report, the maximum allowed internal access time will be called the deadline (for patients to receive treatment). In this thesis, we consider the urgency category of the surgeries that are executed in a Dutch hospital, i.e., urgent and elective patients. Urgent patients are distinguished based on their deadlines. The deadlines of each category is given in the table below.

Urgency class	Deadline to be treated upon arrival
Elective	1 month
24-hour	24 hours
6-hour	6 hours
30-minute	30 minutes

Table 1.1: Patient urgency classes

The data analysis from a large Dutch teaching hospital shows that only a small percentage of urgent surgeries need to be performed within 30 minutes upon their arrival and most of the urgent cases can be delayed for either 6 or 24 hours.

We should note that the patients who need urgent surgeries have adverse conditions and should be treated soon in order to minimize the fatality risk. A study has shown that surgical delay for urgent patients and their mortality risks have close correlation [3]. In The Ottawa Hospital (Canada), McIsaac et al. [3] present that mortality risk for the delayed urgent patients is around 5%, compared to a 3.5% risk of death for those who receive the treatment in time. The study also presents that on average, the delays in urgent surgeries result in a longer length of stay for the patients and a larger operational cost for the hospital than when they are scheduled in time. This implies that finding the optimal schedule for urgent surgeries is pivotal for both the patients and the hospitals.

On the other hand, an operating room is one of the most expensive and scarce resources in the hospital [4], [5]. For this reason, each hospital wants to maximize their OR utilization. Consequently, during office hours, the ORs are booked for elective surgeries to a large extent. In addition to that, in most of the hospitals, some ORs are open outside the office hours, where surgical teams are on standby. In this out-of-office-hours time, the available ORs are dedicated for the emergency patients that need to be treated within 30 minutes upon their arrivals. As the number of these emergency patients is low, the standby resources in the overtime are not fully utilized. The unused standby resources incur an extra cost. To minimize this cost, we can increase the OR utilization by using the standby resources to perform 6-hour and 24-hour surgeries. However, a smart way to schedule the 6-hour and 24-hour surgeries is required, so that the 30-minute patients can be operated in time.

In scheduling urgent surgeries, every time a patient arrives, the OR coordinator has to decide when it should be performed. We call the OR that is dedicated to do the non-elective patients the dedicated OR, and the ones to perform elective surgeries the elective OR. While assigning a 6-hour or 24-hour patient to the dedicated OR can result in a delay in the 30-minute case, allocating this patient to an elective OR during the office hours may result in some cancellations of the elective surgeries. Hence, a hospital needs to schedule urgent patients carefully to treat the patients in time without resulting in excessive elective cancellations.

The existing literature in operations research on scheduling urgent surgeries are scarce. Cardoen et al. [5] show that there are only 20 papers that discuss about non-elective patients (urgent, semi-urgent, and emergency patients). Guerriero and Guido [6] mention in their survey on operating room management (in operations research framework) that scheduling of urgent cases upon their arrivals is interesting for further research. The closest work related to our research is on planning and scheduling semi-urgent patients by Zonderland et al. [7].

In the next section, the questions addressed in this research are presented. Then, the research methodology is given afterwards.

1.1 Research Questions

The goal of this research is to schedule urgent surgeries (6-hour and 24-hour patients) such that the patients receive their treatment in time and OR utilization is maximized. Therefore, the aim is to investigate:

How should urgent surgeries be scheduled in time such that they are not often rejected while also maximizing the OR utilization?

Note that the urgent patients that we consider on this research are the 6-hour and 24-hour patients. Next, by considering the research goal above, we first need to estimate the capacity level that is dedicated for urgent patients. This step is included in the planning of urgent surgeries. Next, in scheduling the urgent surgeries, we only look at the dedicated capacity that we have determined. This implies that the urgent surgeries schedule is independent of the elective patients schedule. The research questions of the scenarios above are formulated as follows:

1. *How much capacity should be dedicated for urgent patients such that the OR utilization is optimal?*
2. *When to schedule an urgent patient in the dedicated capacity such that they are treated in a timely manner?*

To answer the research questions above, the research methodology given in the next section is followed.

1.2 Research Methodology

Initially, a literature study is conducted to gain knowledge related to the focus of the research. The studied literature is related to urgent patient planning and scheduling, for example, reserving capacity using queueing theory approach; scheduling emergency, urgent, or semi-urgent surgeries using Integer Linear Programming (ILP) or Markov decision processes (MDP) models. This step gives us a better viewpoint of the possible approaches that can be used to answer our research questions. Then, by looking at our problems, we build two queueing models to determine the dedicated capacity for urgent surgeries and MDP-based models to schedule them. After proposing the solution methods, we conduct numerical experiments which are inspired by the cases in Jeroen Bosch Ziekenhuis (JBZ), Den Bosch, the Netherlands and RSUP dr. Sardjito (Hospital), Yogyakarta, Indonesia. The Hospital in Indonesia: RSUP dr.Sarjito is further referred as RSUP Sardjito.

1.3 Thesis Outline

The organization of this report is as follows. In Chapter 1, the background of the research as well as the research questions have been presented. In Chapter 2, we give an overview of the related literature studied. Chapter 3 describes $M/M/1/K$ and $M/M/1$ with priority queueuing models to estimate the dedicated capacity level needed to treat urgent patients. Chapter 4 describes the Markov decision processes (MDP) based models to schedule urgent surgeries. In Chapter 5, we present the results of the numerical experiments that are inspired by the cases in the two hospitals: JBZ and RSUP dr.Sardjito using the MDP models. Finally, in Chapter 6 we give the conclusions of our research and the future work.

Chapter 2

Literature review

The concern on healthcare logistics management has increased for the past twenty years. With the growing number of research publications in this field, Cardoen et al. [5] summarize the studies on operating room planning and scheduling in a literature review. More detailed taxonomy classification in healthcare management is done by Hulshof et al. [8].

2.1 Decision Making in Operating Management

A survey on the operational research on operating management by Guerriero and Guido [6] categorizes the literature based on hierarchical decision levels, i.e., strategic, tactical, followed by operational level, which is illustrated by the diagram below.

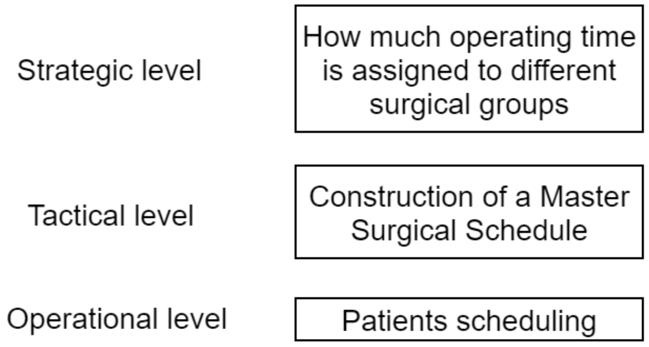
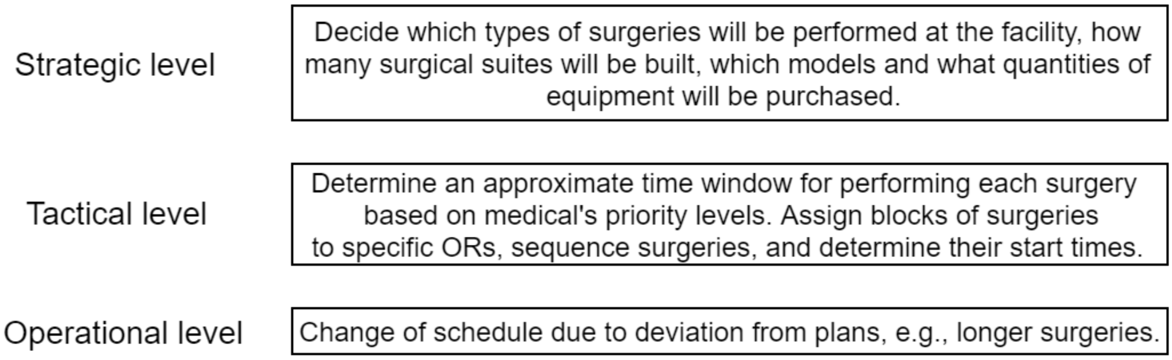

Strategic level
How much operating time is assigned to different surgical groups
Tactical level
Construction of a Master Surgical Schedule
Operational level
Patients scheduling

Figure 2.1: Decision levels in planning and scheduling operating room

Further, Gupta [9] mentions elaborated steps in each decision level regarding elective surgeries scheduling, shown in Figure 2.2. Gupta [9] mentions two ways of scheduling elective surgeries, i.e., block- and open scheduling. In block-scheduling, surgeons are assigned blocks of operating room (OR) time in a periodic schedule. Meanwhile, open scheduling allows surgeons to request for OR time. After that, an OR schedule is constructed before the day of the surgery. This step is included in the tactical level of the decision making process.

Regarding the operational level, Hans et al. [10] split it into off-line and on-line operational level. The off-line operational level deals with constructing robust elective surgical schedules, while the online one takes care of the real-time rescheduling due to unpredictable events, for example, emergency and urgent cases arrivals [10]. As we deal with the online scheduling for urgent surgeries, the focus of this research will be on the tactical level, namely determining the time allocation for urgent surgeries, and online operational level, i.e., scheduling urgent surgeries.



Strategic level

Decide which types of surgeries will be performed at the facility, how many surgical suites will be built, which models and what quantities of equipment will be purchased.

Tactical level

Determine an approximate time window for performing each surgery based on medical's priority levels. Assign blocks of surgeries to specific ORs, sequence surgeries, and determine their start times.

Operational level

Change of schedule due to deviation from plans, e.g., longer surgeries.

Figure 2.2: Decision levels in elective surgery planning and scheduling

From the literature review by Cardoen et al. [5] and Hulshof et al. [8], we see that in the operating room planning and scheduling the most frequently used technique is mathematical programming, followed by simulation, heuristics, Markov processes, and queueing theory. The scenarios where each technique is used will be explained in this chapter. In Section 2.2, we present the literature on the mathematical techniques that are used in surgeries planning and scheduling.

2.2 Operating Room Planning and Scheduling

In this section, we consider the surgery classification given in Table 1.1. Recall that 24-hour, 6-hour, and 30-minute patients fall into urgent patients category that in most of the literature is also called by non-elective surgery. However, the literature on non-elective surgery is scarce [1].

Further, as indicated by the decision levels in Figure 2.1 and 2.2, operating room planning falls into the tactical level and operating room scheduling falls into the operational level. Operating room planning includes the allocation of the resources, such as OR time, staff and beds. In the first subsection, we address the approaches used to allocate time for urgent patients.

2.2.1 Time Allocation for Urgent Surgeries

The surgeries can be categorized either by the medical specialty (the type of procedures needed) or by the urgency level. Several hospitals allocate OR time based on the surgeons' medical specialty, while others decide the time allocation based on the patients' urgency levels. In this report, we consider the time allocation based on patients' urgency levels.

Van Riet and Demeulemeester [1] illustrated three policies in the research which are used to handle non-elective surgeries, i.e., dedicated, flexible, and hybrid policy. However, as hybrid policy is not widely studied, we do not discuss it here. To make the first two policies clearer, the authors demonstrate the idea in the figure below.

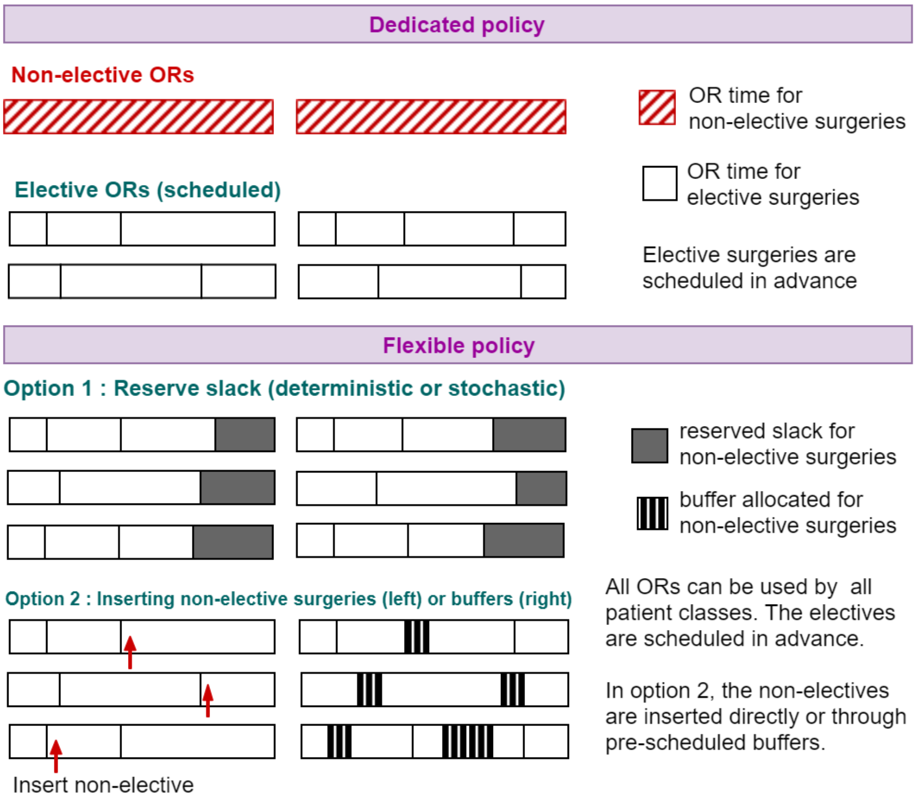

Dedicated policy
Non-elective ORs
OR time for non-elective surgeries
Elective ORs (scheduled)
OR time for elective surgeries
Elective surgeries are scheduled in advance
Flexible policy
Option 1 : Reserve slack (deterministic or stochastic)
reserved slack for non-elective surgeries
buffer allocated for non-elective surgeries
Option 2 : Inserting non-elective surgeries (left) or buffers (right)
All ORs can be used by all patient classes. The electives are scheduled in advance.
In option 2, the non-electives are inserted directly or through pre-scheduled buffers.
Insert non-elective

Figure 2.3: Two policies in handling urgent surgeries [1]

About the two policies in Figure 2.3, Ferrand et al. [11] make a comparison on the waiting time of both elective and non-elective (emergent) cases as well as the overtime. In the dedicated policy, the waiting time and overtime on elective cases are lower than those in the flexible policy as their schedule is not disturbed by the non-elective cases. On the contrary, the waiting time for the non-elective case in the dedicated policy is higher than in the flexible one. This is because the non-elective surgeries cannot disrupt the elective OR time and have to wait until the procedure in the dedicated OR is finished.

With regard to the use of dedicated policy in treating the non-elective patients, Zonderland et al. [7] focus on the semi-urgent patients that are distinguished based on their maximum waiting time: 1-week and 2-week semi urgent patients. The authors use a queueing theory framework to evaluate the reserved OR time for the semi-urgent patients in the long term. However, due to unpredictability of the semi-urgent patients, reserving OR time may yield in unused OR time due to overbooking reserved OR time and elective cancellations due to semi-urgent patients. To tackle this problem, another queueing model is proposed to balance the number of elective cancellations and unused reserved OR time. Using a different approach, Gerchak et al. [12] take into account open scheduling. They look at the case where the OR capacity utilization by the elective and emergency cases is uncertain. In this paper, new requests to book the OR time arrive every day. A stochastic dynamic programme is constructed to calculate the amount of OR time that should be reserved for the elective cases such that the hospital can perform possible emergent surgery.

As for the flexible policy, Van Riet and Demeulemeester [1] explain that the amount of slack for option 1 of the flexible policy (shown in Figure 2.3) can be determined using a queueing theory framework as done by Zonderland et al. [7]. Meanwhile, van Essen et al. [13] use BIM optimization to insert emergency surgery in between the elective surgeries. This idea is similar to the left part of option 2 in the flexible policy.

By considering add-on surgeries as the surgeries that need to be performed on the arrival day (emergency and urgent cases), Zhou and Dexter [14] predict the upper bound of the total add-on surgical duration. By assuming that the case duration follows a log-normal distribution, they could predict the maximum slack to allocate for add-on surgeries. After analyzing the possible way of allocating time for urgent surgeries, we examine the tactical level in decision making, namely scheduling surgery.

2.2.2 Scheduling Urgent Surgeries

One of the approaches in scheduling elective patients is by using a Master Surgery Schedule (MSS). In the decision making, this part is categorized in off-line tactical level. There are many studies with various mathematical techniques employed to build an optimal MSS. Among the literature on MSS are Beliën et al. [15] who employ Mixed Integer Programming (MIP) by taking the daily mean and variance of the bed occupancy. The authors also aim to centralized the surgeons that belong in one group (one medical specialty) in one operating room, as well as making the schedule as repetitive as possible in order to form a cyclic schedule. The three objectives all together build a large MIP that leads to the use of a heuristic method to obtain a solution. The heuristic method results in a trade-off between bed occupancy and operating room centralization based on surgeons' medical specialty.

Focusing on urgent surgeries, Dexter et al. [16] propose ILP models to sequence urgent surgeries by considering the medical procedure needed, estimated surgery duration and the deadline. The estimated surgery duration is obtained from historical data. Also, the urgency level of the urgent patient is determined from the evidence in medical literature. There are three objectives addressed, i.e., minimizing the average waiting time of the urgent patient, sequencing urgent patients when the First-Come, First-Served (FCFS) rule is used, and sequencing patients based on the urgency level (medical priority). These three objectives are studied separately with only one constraint, namely the starting time of the surgery should not exceed the deadline.

As the literature on scheduling urgent surgery is scarce, we explore papers on semi-urgent and emergency cases, as both are included in the non-elective patient category. Concerning emergency patients, some hospitals enable them to be performed in the elective ORs which causes elective surgery cancellations. Employing this pollicy, Erdem et al. [17] propose a mixed integer linear programming (MILP) model to reschedule the cancelled elective surgeries.

Our work is related to the paper of Zonderland et al. [7] where 1- and 2-week semi-urgent patients are the focus. After determining the reserved OR time for semi-urgent patients using queuing models, Zonderland et al. [7] develop a model based on Markov decision processes (MDP) to schedule the elective and semi-urgent cases. In this model, the elective case can be cancelled to perform semi-urgent surgeries. The cancelled electives then become 1-week semi urgent surgeries. The authors apply the models in the dataset from a neurosurgery department of a Dutch academic hospital. In our work, the MDP approach is used in our models to schedule the urgent patients.

Chapter 3

Dedicating capacity for urgent surgeries

At the strategic level of operating room management, the operating room (OR) manager needs to determine the dedicated capacity level for the urgent patients in the long run. In this chapter, we use two queuing models for this goal. Based on the results of these models, the OR manager can decide the dedicated capacity level to handle the arriving urgent patients by considering the OR utilization and patients' waiting times. In the next section we present the assumptions employed in the queueing models.

3.1 Assumptions

Recall that in this research, we consider the 6-hour and 24-hour urgent patients. We describe the arrival process of the urgent patients in the following assumption.

Assumption 3.1.1. *Each type of urgent patient arrives according to a Poisson process. The interarrival times are exponentially distributed and independent of each other.*

Hence, the arrival process of the urgent patients is a Poisson process. Regarding the OR day and the dedicated OR, we have the following assumptions.

Assumption 3.1.2. *OR day is divided into 4 shifts, where in each shift, the number of patients that can be treated is the same, i.e., s patients per shift.*

Assumption 3.1.3. *There is one dedicated OR to perform urgent surgeries.*

In this report, the server is the dedicated OR and the service rate is the number of surgeries performed in each OR per shift.

Assumption 3.1.4. *The service rate is exponentially distributed and independent of the arrival process.*

In the next part of the report, we use the Kendall's notation on the queueing model, i.e., $A/B/X/Y/Z$, where A indicates distribution of the interarrival time, B shows the distribution of the service time, X is the number of the parallel servers, Y is the maximum capacity of the queue, and Z is the service discipline. For exponentially distributed interarrival and service times it is denoted by M for memoryless in the Kendall's notation. Some alternatives of the service discipline are first come, first served (FCFS); last come, first served (LCFS); priorities.

3.2 Queueing models to dedicate capacity for urgent patients

Using the assumptions given in the previous section, in this section, we look at two queueing models and conduct a case study that is inspired by the cases in Jeroen Bosch Ziekenhuis(JBZ), Den Bosch and RSUP Sardjito, Yogyakarta, Indonesia. First, we look at the case where the hospital set a maximum number of urgent patients that can be admitted to the system. Denote K patients as the maximum number of urgent patients in the system. For this scenario, we look at the $M/M/1/K$ queueing model.

3.2.1 $M/M/1/K$ queueing model

In this model, the arriving patients are rejected when there are K patients in the system. From Assumptions 3.1.2 and 3.1.3, denoting the dedicated capacity level by s patients per shift, we have $\mu = s$ in this model. For this case, μ is the service rate. Next, the occupation rate of this queueing model is given by:

$$\rho = \frac{\lambda}{\mu}, \tag{3.1}$$

where λ is the total arrival rates of the urgent patients per shift, μ is the service rate (number of urgent surgeries performed) per shift. Hence, have the probability of having n patients in the system as follows :

$$p_n = \frac{\lambda^n}{\mu^n} p_0 \,, \qquad\qquad 0 \le n \le K. \tag{3.2}$$

Using normalization, we obtain:

$$p_0 = \left(\sum_{n=0}^{K} \frac{\lambda^n}{\mu^n} \right)^{-1} = \begin{cases} \frac{1}{K+1}, & \rho = 1 \\ \frac{1-\rho}{1-\rho^{K+1}}, & \rho \neq 1. \end{cases} \tag{3.3}$$

The expected queue length is given by [18]:

$$E(L_q) = p_0 \frac{\lambda\, \rho}{\mu\, (1-\rho)^2} [1 - \rho^K - (1-\rho)(K)\rho^{K-1}]. \tag{3.4}$$

We should keep in mind that in $M/M/1/K$ queueing model the capacity of the system is finite and bounded by K. Thus, a fraction of arrivals cannot enter the system (denoted by p_K), because they arrive when the system is full. In this report, the arrivals that can enter the system are called effective arrivals. The effective arrival, denoted by λ_{eff}, takes place when there are less than K patients in the system. Using PASTA property (Poisson arrivals see time averages), the effective arrival rate seen by the servers is $\lambda_{eff} = \lambda(1-p_K)$. Using λ_{eff}, we obtain the expected number of patients in the system as follows:

$$\begin{aligned} E(L) &= E(L_q) + \frac{\lambda_{eff}}{\mu} \\ &= E(L_q) + \frac{\lambda(1-p_K)}{\mu}. \end{aligned} \tag{3.5}$$

In the formula above, $\frac{\lambda_{eff}}{\mu}$ shows the number of patient that is currently performed. This implies $\frac{\lambda_{eff}}{\mu} < 1$ because the average number of surgeries performed cannot exceed the available OR capacity. Next, using Little's formula we obtain the following expected time a patient spends in the system:

$$\begin{aligned} E(W) &= \frac{E(L)}{\lambda_{eff}} \\ &= \frac{E(L)}{\lambda(1-p_K)}, \end{aligned} \tag{3.6}$$

and the expected waiting time :

$$\begin{aligned} E(W_q) &= \frac{E(L_q)}{\lambda_{eff}} \\ &= \frac{E(L_q)}{\lambda(1-p_K)}. \end{aligned} \tag{3.7}$$

More detailed and general formulas of this model can be found in Shortle et al. [18]. In $M/M/1/K$ queueing model the patients are treated according to First Come First Served (FCFS) rule. This means that the 6-hour patients do not have priority over the 24-hour patients. Next, we compare this model to the $M/M/1$ queueing model with priority rule in giving the treatment.

3.2.2 $M/M/1$ with priority rule queueing model

In this model, we look at two types of patients: 6-hour and 24-hour patients. The 6-hour patients have priority over the 24-hour patients. We employ non-preemptive priority where the 6-hour patients may not interrupt the surgery time (service time) of the 24-hour patients.

Employing Assumption 3.1.1, the 6-hour and 24-hour patients arrive according to a Poisson process with rate λ_6 and λ_{24}, respectively. The service rate for both patient types are identical, i.e., μ patients per shift. The occupation rate of the system due to the 6-hour patients is:

$$\rho_6 = \frac{\lambda_6}{\mu}. \tag{3.8}$$

Next, due to the 24-hour, the occupation rate is as follows:

$$\rho_{24} = \frac{\lambda_{24}}{\mu}. \tag{3.9}$$

From Equations (3.8) and (3.9), the occupation rate of the system is $\rho = \rho_6 + \rho_{24}$. The system is stable when $\rho < 1$. For the queueing analysis, we have the expected time that a patient spends in the system (waiting in the queue and being treated) formulated below [19]:

$$E(W_6) = \frac{(1+\rho_{24})/\mu}{1-\rho_6}. \tag{3.10}$$

Using the Little's law, we obtain the expected number of 6-hour patients in the system, formulated by:

$$E(L_6) = \frac{(1+\rho_{24})/\rho_6}{1-\rho_6}. \tag{3.11}$$

The expected number of 24-hour patients in the system in the system is:

$$E(L_{24}) = \frac{(1-\rho_6(1-\rho_6-\rho_{24}))\rho_{24}}{(1-\rho_6)(1-\rho_6-\rho_{24})}. \tag{3.12}$$

Using the Little's law, we obtain expected time the 24-hour patients spend in the system, formulated by:

$$E(W_{24}) = \frac{E(L_{24})}{\lambda_{24}} = \frac{(1-\rho_6(1-\rho_6-\rho_{24}))/\mu}{(1-\rho_6)(1-\rho_6-\rho_{24})}. \tag{3.13}$$

For the expected time 6-hour patients spend in the queue, we have the following formula:

$$E(W_{q6}) = E(W_6) - \frac{1}{\mu} = \frac{1}{\mu}\left(\frac{1+\rho_{24}}{1-\rho_6} - 1\right). \tag{3.14}$$

Next, using Little's law the expected number of 6-hour patients spend in the queue is:

$$E(L_{q6}) = \lambda_6 \,.\, E(W_{q6}) = \rho_6\left(\frac{1+\rho_{24}}{1-\rho_6} - 1\right). \tag{3.15}$$

Using the similar way, we obtain the expected time 24-hour patients spend in the queue, we have the following formula:

$$E(W_{q24}) = E(W_{24}) - \frac{1}{\mu} = \frac{1}{\mu}\left(\frac{(1-\rho_6(1-\rho_6-\rho_{24}))}{(1-\rho_6)(1-\rho_6-\rho_{24})} - 1\right). \tag{3.16}$$

Next, according to Little's law, we obtain:

$$E(L_{q24}) = \lambda_{24}\,.\,E(W_{q24}) = \rho_{24}\left(\frac{(1-\rho_6(1-\rho_6-\rho_{24}))}{(1-\rho_6)(1-\rho_6-\rho_{24})} - 1\right). \tag{3.17}$$

In the next section, we conduct a case study for a various dedicated capacity levels using the queueing models in Section 3.2.

3.3 Case study of queueing models to dedicate capacity for urgent patients

In this section, we present the results of the case study on the $M/M/1/4s$ and $M/M/1$ with priority rule queueing models conducted using Qtsplus 3.0 on a Lenovo ThinkPad E550 with Intel(R) Core(TM) i5-5200 CPU @ 2.20GHz processor (8GB RAM). We use the parameters that are inspired by the cases in Jeroen Bosch Ziekenhuis (JBZ), Den Bosch in 2017 and RSUP Sardjito (Hospital), Yogyakarta, Indonesia in May-August 2019. For both hospitals, we look at the data where each OR day is divided into 4 shifts, each of 6-hour length. We calculate the arrival rate of each patient type per shift. The arrival rates of the 6-hour and 24-hour patients at JBZ in 2017 are given in the following table.

	6-hour	**24-hour**
$\bar{\lambda}$	0.54	1.52
λ_h	0.9	3.03

*$\bar{\lambda}$= average arrival rates
λ_h= highest arrival rates
(no.of patients/shift)

Table 3.1: Arrival rates of the 6-hour and 24-hour patients in the JBZ dataset

Considering the arrival rates in the table above, for JBZ case, the sensitivity analysis is conducted using the following arrival rates.

No	Arr. rate (patients/shift)
1	6-hour patients : 0.54 24-hour patients : 1.2
2	6-hour patients : 0.54 24-hour patients : 1.52
3	6-hour patients : 0.9 24-hour patients : 3.03

*Arr. rate = arrival rate of each patient type per shift (no.of patients/shift);
s = the maximum number of surgeries that can be performed in the dedicated capacity.

Table 3.2: The various arrival rates for case study on queueing models based on JBZ dataset

In the next parts, we conduct numerical experiments for the cases above using $M/M/1/4s$ and $M/M/1$ with priority rule queueing models.

Whereas in RSUP Sardjito, only arrivals of the 24-hour patients take place. The arrival rate is 0.817 patients per shift in average, where 1.82 patients per shift is the highest arrival rate. For this case, we only conduct a case study using $M/M/1/4s$ queueing model as there is only one type patient in the system.

3.3.1 Case study on $M/M/1/4s$ queueing model

In $M/M/1/4s$ queueing model, the dedicated OR has a capacity of $\mu = s$ surgeries per shift. The maximum number of patients in the system, $4s,$ is obtained by considering that the 24-hour patients can afford to wait up to 4 shifts upon their arrivals. The case study is conducted for the arrival rates given in Table 5.4. For each case, we look at the server utilization compared to the expected number of rejected patients, queue length and waiting time in the queue.

For the first case, where $\lambda_6 = 0.54$ and $\lambda_{24} = 1.2$, the sensitivity analysis is performed by taking $s = 2$ to $s = 8.$ The chart in Figure 3.1 shows the server utilization along with the expected number of rejected patients, queue length and waiting time length in each shift. We can observe that the higher server utilization yields in the larger queue length, which implies a longer waiting time and more rejected patients. For example, take the case where $s = 2$.The server utilization is around 0.8, while the expected queue length is around 2.3 patients per shift ($\approx$ 8 patients a day) and the expected waiting time is 1.4 shifts where in average 0.1 patients are rejected per shift. The maximum capacity that we test is not larger than $s = 8$, because we can see that from $s = 3$ to $s = 8$ no patients are rejected (all arrivals can be admitted).

For the total arrival rates of $\lambda = 2.06$ and $\lambda = 3.93,$ the charts in Figures 3.2 and

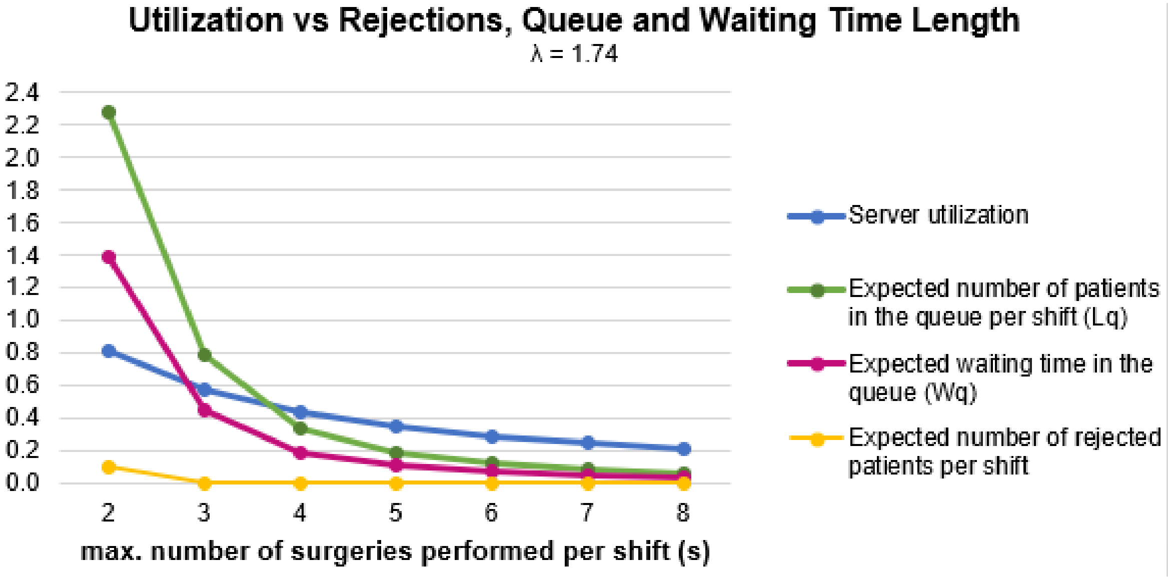


Figure 3.1: Sensitivity analysis for $\lambda_6 = 0.54$ and $\lambda_{24} = 1.2$

3.3 respectively illustrate the server utilization along with the expected number of rejected patients, queue length, and waiting time length in each shift for different capacity s. Similar with the first case, in these two cases, the higher utilization yields in larger expected queue length and longer expected waiting time. The highest

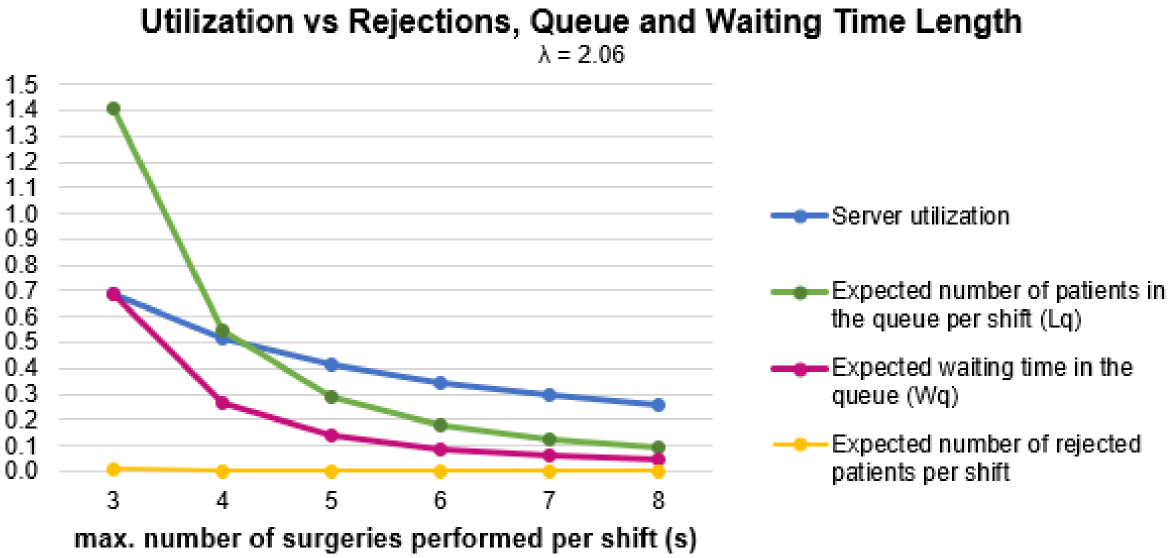


Figure 3.2: Sensitivity analysis for $\lambda_6 = 0.54$ and $\lambda_{24} = 1.52$

server utilization of arund 0.9 is reached for $s = 4$. However, for this case, the expected number of patients waiting in a shift is 6.6 and in average we reject around 0.3 patient in a shift (≈ 2 patients in a day). The expected queue length drops to 2.7 patients a shift when $s = 5$. This affects the expected rejections and waiting length to be 0 patients per shift and around 0.9 shift, respectively.

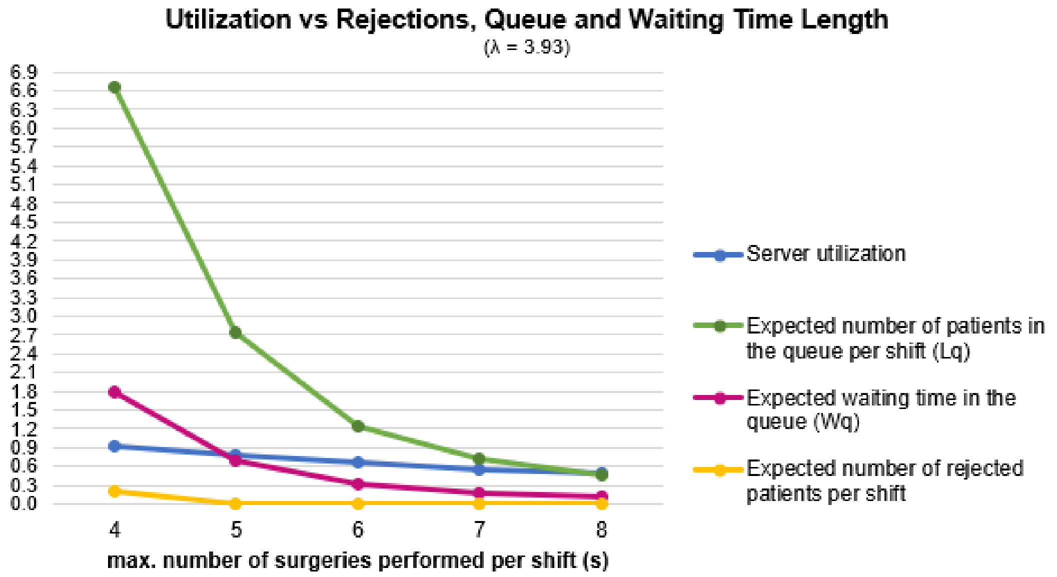


Figure 3.3: Sensitivity analysis for $\lambda_6 = 0.9$ and $\lambda_{24} = 3.03$

Next, the results of the case study for the dataset from RSUP Sardjito are presented in Figures 3.4 and 3.5 for $\lambda = 0.817$ and $\lambda = 1.82$, respectively. Same as the previous results on JBZ dataset, the higher utilization results in larger mean queue length and longer mean waiting time.

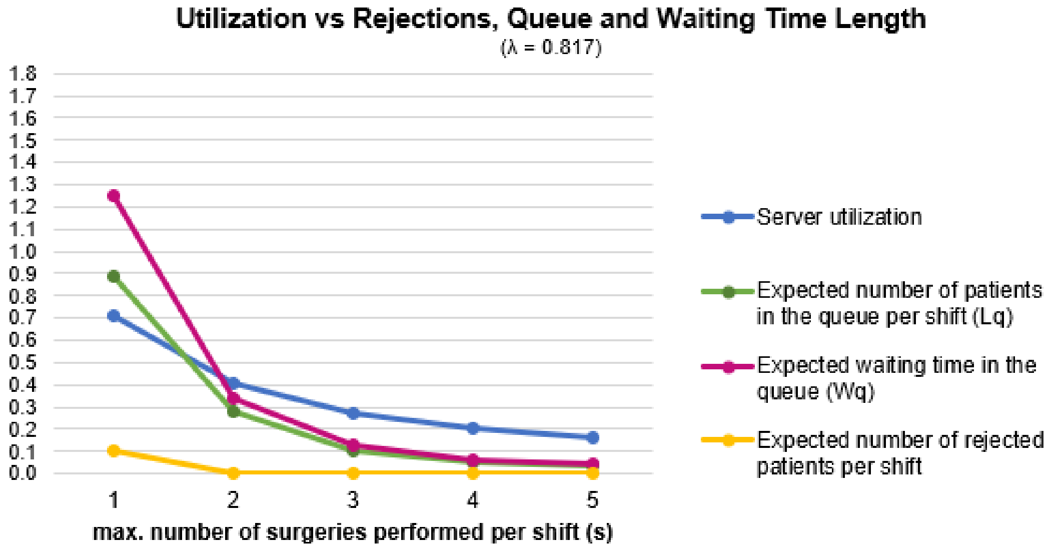


Figure 3.4: Sensitivity analysis for $\lambda_{24} = 0.817$

Finally, it depends on which parameters hospitals focus on to decide the suitable dedicated capacity. In the next part, we use the parameters from JBZ dataset to perform sensitivity analysis using $M/M/1$ with priority queueing model.

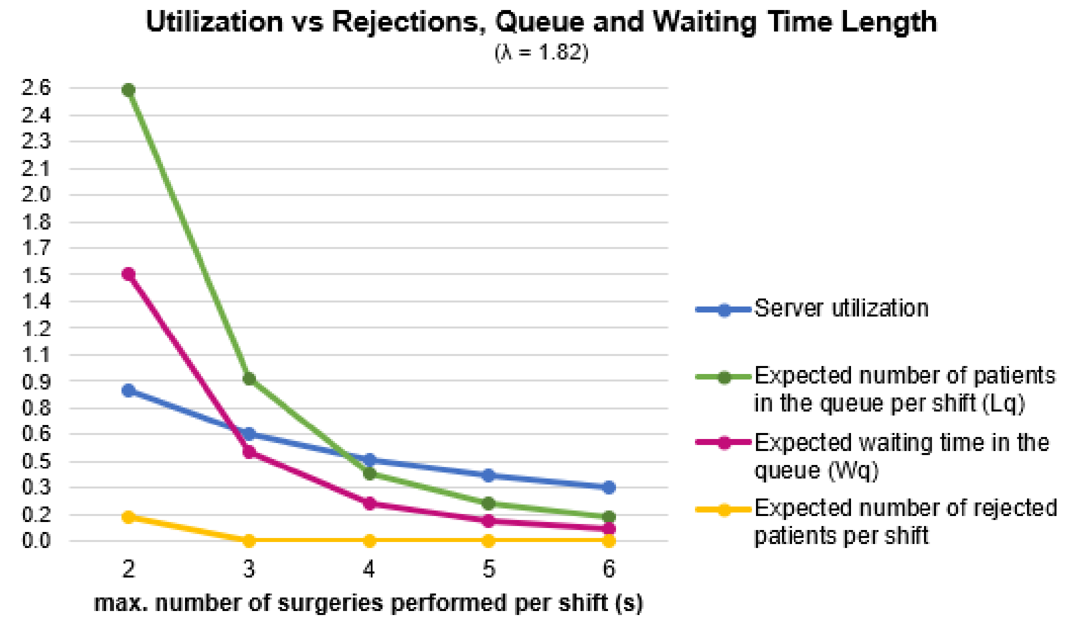


Figure 3.5: Sensitivity analysis for $\lambda_{24} = 1.82$

3.3.2 Case study on $M/M/1$ with priority rule queueing model

In $M/M/1$ with priority queueing model, the queue has one dedicated OR (server) with service rate of μ patients per shift. By looking at Assumption 3.1.2, we have $\mu = s$ patients per shift. The two type patients are treated using non-preemptive priority rule, where the 6-hour patients have priority over the 24-hour patients.

First, we look at the server utilization and expected waiting time for 6-hour and 24-hour patients, where $\lambda = 1.74$ and $s = 2$ to $s = 8$ given in Figure 3.6. We also observe the server utilization and expected queue length for 6-hour and 24-hour patients for this case in Figure 3.7. The sensitivity analysis results using other parameters in Table 5.4 for this model are given in the Appendix A.

Recall that in the data from RSUP Sardjito there is no 6-hour patient arrivals, i.e., $\lambda_6 = 0$. Hence, the $M/M/1$ with priority queueing model is not suitable for this data as we only have one patient type: 24-hour patients.

Last, we compare the mean waiting time in $M/M/1/4s$ and $M/M/1$ priority queueing models for the parameters in Table 5.4. The comparison of these models when $\lambda_6 = 0.54$, $\lambda_{24} = 1.2$ is given in the following chart. We can see from the chart in Figure 3.8 that for $s = 2$ in the $M/M/1$ priority queueing model the waiting time of the 6-hour patients improves by around 0.5 shift. However, due to the priority rule, the waiting time of the 24-hour patients gets worse by 3 shifts in average. The expected

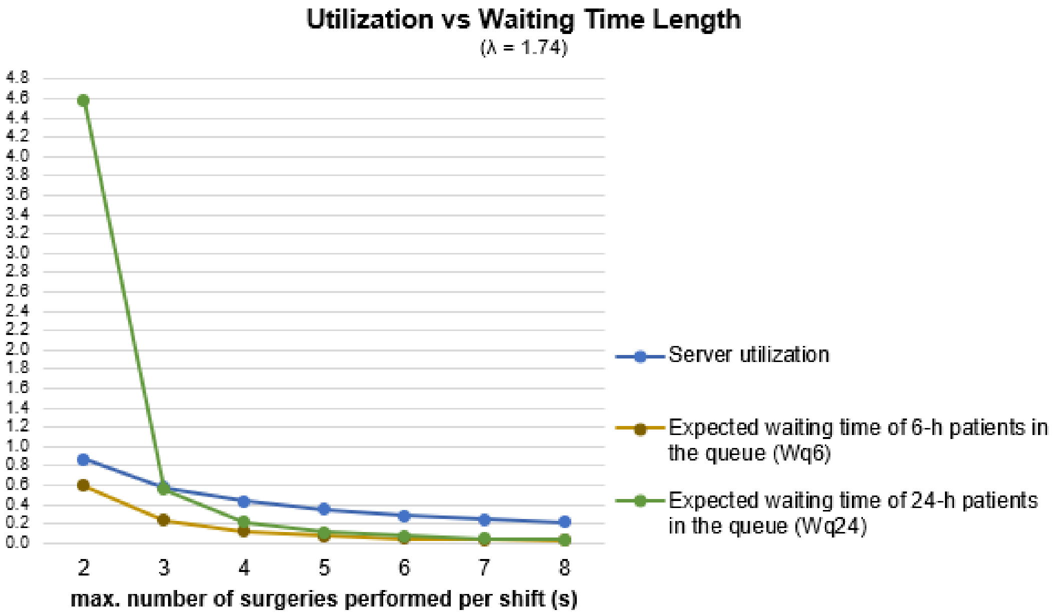


Figure 3.6: Sensitivity analysis on the mean waiting time for $\lambda_6 = 0.54$; $\lambda_{24} = 1.2$ using priority rule

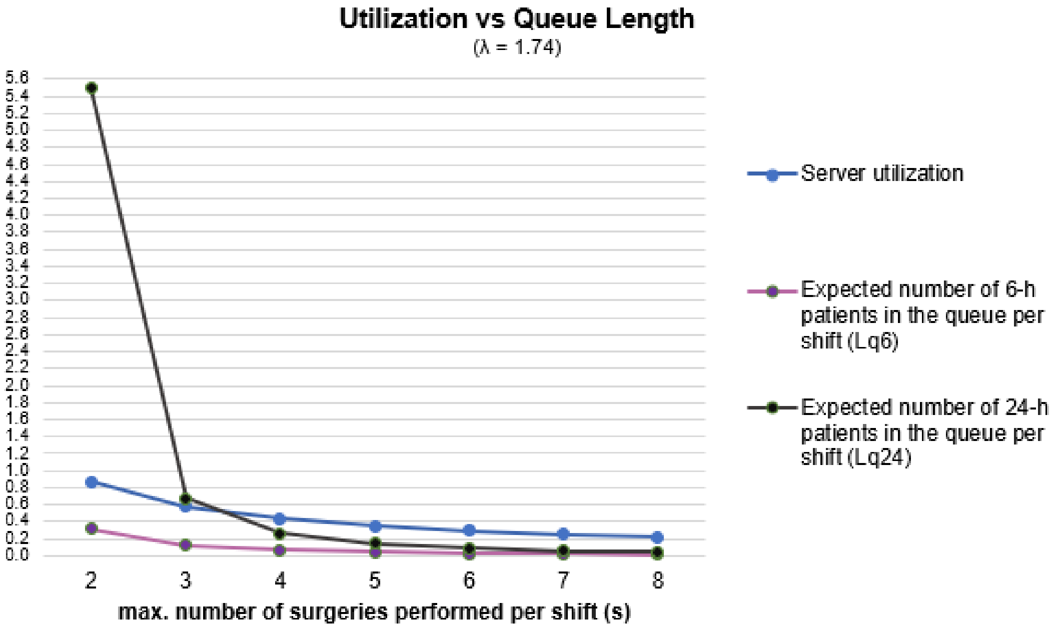


Figure 3.7: Sensitivity analysis on the mean queue length $\lambda = 1.74$ using priority rule

waiting times of both patients drop when the hospital has a capacity of at least $s = 3$. For $s = 4$, the mean waiting times of both patient types in the two queueing models are close, where the gaps are getting smaller as s gets larger.

Next, we see the average waiting times for $\lambda_6 = 0.54$ and $\lambda_{24} = 1.52$ in the two queueing models.

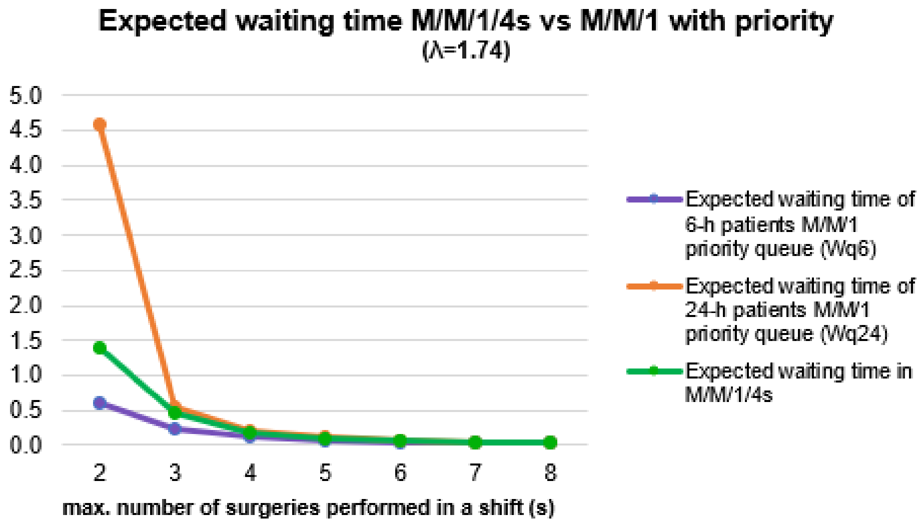


Figure 3.8: Expected waiting time for $\lambda_6 = 0.54$ and $\lambda_{24} = 1.2$ in $M/M/1/4s$ and $M/M/1$ priority queueing models

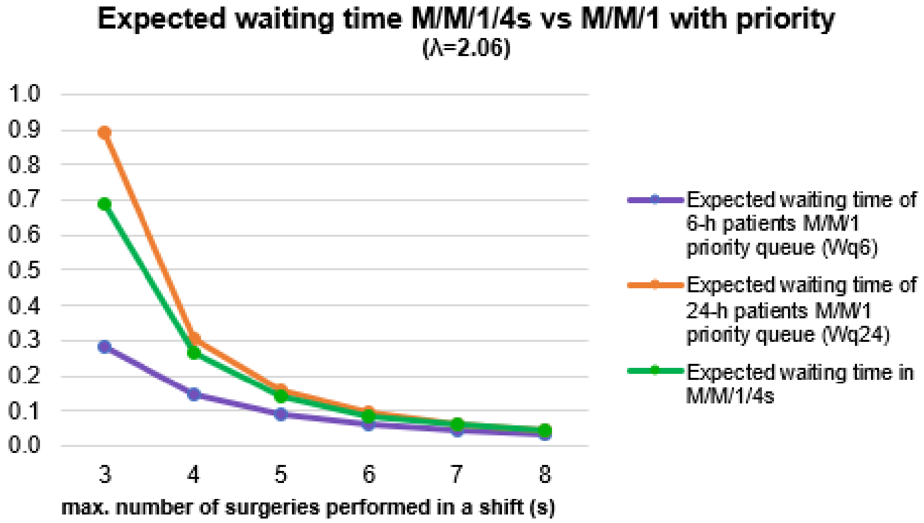


Figure 3.9: Expected waiting time for $\lambda_6 = 0.54$ and $\lambda_{24} = 1.52$ in $M/M/1/4s$ and $M/M/1$ priority queueing models

From the results in the two queueing models in the three last figures, we can conclude that high utilization results in large queue length and waiting time (other figures are attached in Appendix A).

Hence, the results from the two queueing models can be used to balance the OR utilization, and patients' waiting times or queue length of each patient type or the number of rejected patients, depending on the hospital's preferences. The queueing model used also depends on the patient arrivals. In the case of RSUP Sardjito where we consider that only 24-hour patients arrive to the system, $M/M/1$ with pri-

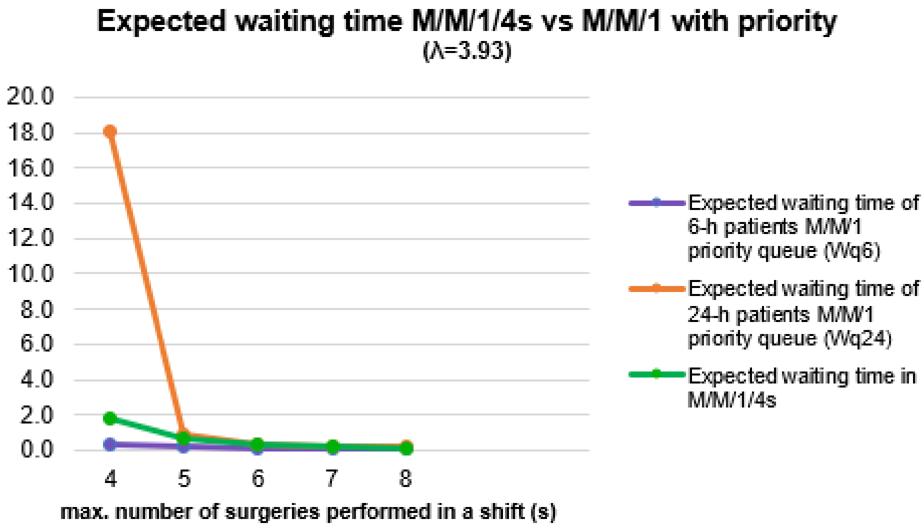


Figure 3.10: Expected waiting time for $\lambda_6 = 0.9$ and $\lambda_{24} = 3.03$ in $M/M/1/4s$ and $M/M/1$ priority queueing models

ority queueing model cannot be used as there is only one type of patient. Thus, for such case, $M/M/1/4s$ queueing model is suitable to determine the dedicated capacity level needed for the urgent surgeries.

Chapter 4

Model for scheduling urgent surgeries independent of elective surgeries

This chapter formulates four models to schedule urgent surgeries independent of the elective patients based on Markov decision processes (MDP) which employs the infinite planning horizon for the hospital that uses dedicated operating rooms to perform urgent surgeries.

This chapter is structured as follows: Section 4.1 describes what Markov decision processes is, Section 4.2 gives the assumptions that are used to formulate the models, Section 4.3 describes the MDP to decide the number of urgent surgeries performed in each part of the day along with the MDP that allows deferring the patients to other resources in Subsection 4.3.1, Section 4.4 explains the MDP to assign a number of urgent surgeries to the appropriate time.

4.1 Markov decision processes (MDP)

In Markov decision processes (MDP) we observe a process at discrete time points in infinite horizon. At each time point, the system is at one of the possible states. Transition probabilities among states only depend on the current state and not the previous states. By considering the transition probabilities among states for each possible action as well as the direct costs corresponding to it, an optimal action is chosen at each point of time. In the subsequent time point the decision maker have to chose the optimal action. The discrete time points where an action should be chosen are called decision epochs. In the proposed models, we decide the optimal action by considering the costs incurred in the current time point as being more important than those incurred in the future. More about MDP can be found in [20].

In the next sections, as we build the model using a Markov decision processes (MDP), the following elements are needed [20]:

1. Decision epochs.
2. States.
3. Actions.
4. Transition probabilities.
5. Direct costs.

In formulating the models, the assumptions used are explained in the next section.

4.2 Assumptions

This section explains the assumptions made to build the models. Recall the urgent patients categories in Table 1.1. Assumption 4.2.1 gives the policy to handle the 30-minute surgeries.

Assumption 4.2.1. *30-minute surgeries are performed within the shift where they arrive.*

Assumption 4.2.1 implies that the 30-minute surgeries are scheduled and hence will not be considered in the models. Assumption 4.2.2 gives the distribution of urgent patient arrivals.

Assumption 4.2.2. *Each type of urgent patient arrives at a shift according to a Poisson process. The arrival process within a shift is independent of the arrival processes in other shifts.*

A Poisson process is a counting process. The definition of a counting process is shown in Definition 4.2.3. The definition of a Poisson process is shown in Definition 4.2.4.

Definition 4.2.3. *[21] A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of events that occur by time t.*

Definition 4.2.4. *[21] The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process having rate $\lambda > 0$ if*

1. $N(0) = 0$
2. *The process has independent increments (the number of events that occur in disjoint time intervals are independent)*

3. *The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all s, $t \geq 0$*

$$\mathbb{P}(N(t+s) - N(s) = n) = e^{\lambda.t}\frac{(\lambda.t)^n}{n!}, \qquad n = 0, 1, 2, \ldots. \tag{4.1}$$

In further part of this report, the Poisson probability above is denoted by

$$Poi(n, \lambda.t).$$

The daily OR time is divided into four shifts: morning, afternoon, evening and night. Assumption 4.2.5 is needed to make the decision for each urgent surgery type.

Assumption 4.2.5. *The 6-hour patients should be scheduled one shift ahead of the arriving shift and the 24-hour patients can wait up to four shifts upon the arriving shift.*

The models are constructed for hospitals that dedicate some capacity for urgent patients. Let s_n denote the maximum number of urgent patients that can be performed in shift n using the dedicated capacity. Assumption 4.2.6 gives the maximum number of surgeries that can be performed in each shift that we take into account in the models.

Assumption 4.2.6. *The maximum number of urgent surgeries that can be performed within each shift is identical, regardless of the type of surgeries. Hence, we have $s_n = s$ for all $n \in \mathbb{N}$.*

In Sections 4.3 and 4.4 we describe the MDP model to schedule the urgent patients by observing the arrivals until the end of each shift where an action should be taken.

4.3 Assigning a number of urgent patients in each shift (Model 1)

In this section we describe the MDP model to schedule urgent patients, where in each shift we determine the number urgent surgeries of each type that are assigned in the next shift.

Decision epochs

Decision epoch is the moment when an action should be taken. In this case, the action is taken at the end of each shift and shown in Figure 4.1.

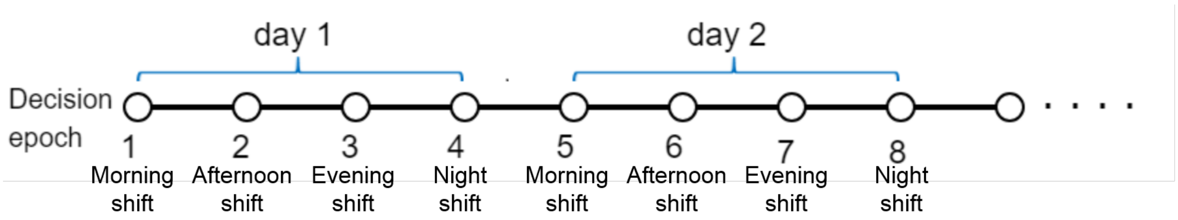


Figure 4.1: Decision epoch

The notation for the decision epoch is given by:

$$\mathbb{N} = \{1, 2, 3, \ldots\},$$
$$n \in \mathbb{N}.$$

State space

The state at the end of shift $n \in \mathbb{N}$ is the number of urgent cases that can wait for a maximum of t more shifts upon its arrival. By employing Assumption 4.2.5, the definition below is given to model the state space.

Definition 4.3.1. *Let a stochastic process* $\{\mathbf{U}_n,\ n = 1, 2, 3, \ldots\}$, *where*

$$\mathbf{U}_n = (U_{1,n}, U_{2,n}, U_{3,n}, U_{4,n}), \forall n \in \mathbb{N}$$

and $U_{t,n}$ *records the number of urgent patients that are allowed to wait for* t *shifts in the system at the end of shift* $n, \forall n \in \mathbb{N},\ t = 1, 2, 3, 4.$

Let u denote the realization of the random variable U in Definition 4.3.1. Hence, the state at the end of shift n is denoted by:

$$\boldsymbol{u_n} = (u_{1,n}\,;\, u_{2,n}\,;\, u_{3,n}\,;\, u_{4,n}).$$

All possible states at decision epoch n build the entire state space $\mathscr{S}$, which is given by:

$$\mathscr{S} = \{S_n\}_{n\in\mathbb{N}},$$

where

$$S_n = \{\boldsymbol{u_n} = (u_{1,n}\,;\, u_{2,n}\,;\, u_{3,n}\,;\, u_{4,n}) \,|\, u_{1,n},\ u_{2,n},\ u_{3,n},\ u_{4,n} = 0, 1, 2, \ldots < \infty\}\ \ \forall n \in \mathbb{N}. \tag{4.2}$$

Boundary of the states

By looking at the definition of the states above and Assumption 4.2.6, the total number of surgeries in the system within each shift should not exceed $4s$, which is formulated in the following equation:

$$\sum_{t=1}^{4} u_{t,n} \leq 4s,\ \forall n \in \mathbb{N}. \tag{4.3}$$

Decisions

The decision in each shift is to determine the number of urgent surgeries that are performed in the next shift. Other than the maximum waiting time of urgent surgeries that is given by Assumption 4.2.5, the decisions should satisfy the conditions below:

1. If some surgeries are not yet assigned, then they can wait for one shift less. For example, if in shift $n \in \mathbb{N}$ the state is given by $\boldsymbol{u_n} = (0, 1, 1, 1)$ and no surgeries are assigned to the next shift, then in the next shift the state is $\boldsymbol{u_{n+1}} = (1, 1, 1, 0)$.
2. The sum of the surgeries that are assigned in each shift should not exceed the dedicated capacity.

The notation for the action is as follows:

$$\begin{aligned} A_{\boldsymbol{u},n} = \{ & \boldsymbol{a_n} = (a_{1,n}, a_{2,n}, a_{3,n}, a_{4,n}) = (\text{do } a_{t,n} \text{ surgery out of } u_{t,n},\ t = 1, 2, 3, 4), \\ & \forall n \in \mathbb{N} \,|\, a_{t,n} \leq u_{t,n}, \sum_{t=1}^{4} a_{t,n} \leq s, \quad t = 1, 2, 3, 4,\ \forall n \in \mathbb{N}\}, \\ A = & \bigcup_{\boldsymbol{u} \in \mathcal{S}, n \in \mathbb{N}} A_{\boldsymbol{u},n}. \end{aligned}$$

Transition probabilities

In formulating the transition probabilities, we take patient arrivals into account. Let random variable $R_{p,n}$ denote the number of type p patients that arrive to the system at shift n, where $p = 6$ and $p = 24$ represent the 6-hour and 24-hour surgeries, respectively. Using Assumption 4.2.2, $R_{p,n}$ follows a Poisson process with an arrival rate of λ_p, where $p = 6, 24,$ for all $n \in \mathbb{N}$.

The evolution of the states given by the diagram in Figure 4.2 is used to construct the transition probability. Also, based on Assumption 4.2.6, the following assumption describes the policy to admit the arriving patients.

Assumption 4.3.2. *First, the maximum number of 6-hour patients are admitted to fill the slot s. Next, the 24-hour patients are admitted such that the total number of patients within a shift does not exceed $4s$. In case the system is full, i.e., there are $4s$ patients in the system, we reject all arrivals from entering the system.*

Let $\bar{R}_{p,n}$ denote the number of type p patient arrivals that are admitted at shift n, where $p = 6, 24$ denoting 6-hour and 24-hour patients, respectively. Next, from Figure 4.2 and Assumption 4.3.2, we admit $\bar{R}_{6,n}$ 6-hour arrivals, where $\bar{R}_{6,n} = \min\{R_{6,n}, s - u_{2,n-1} + a_{2,n-1}\}$. Hence, $u_{1,n} = \min\{u_{2,n-1} - a_{2,n-1} + R_{6,n}, s\}$. Next,

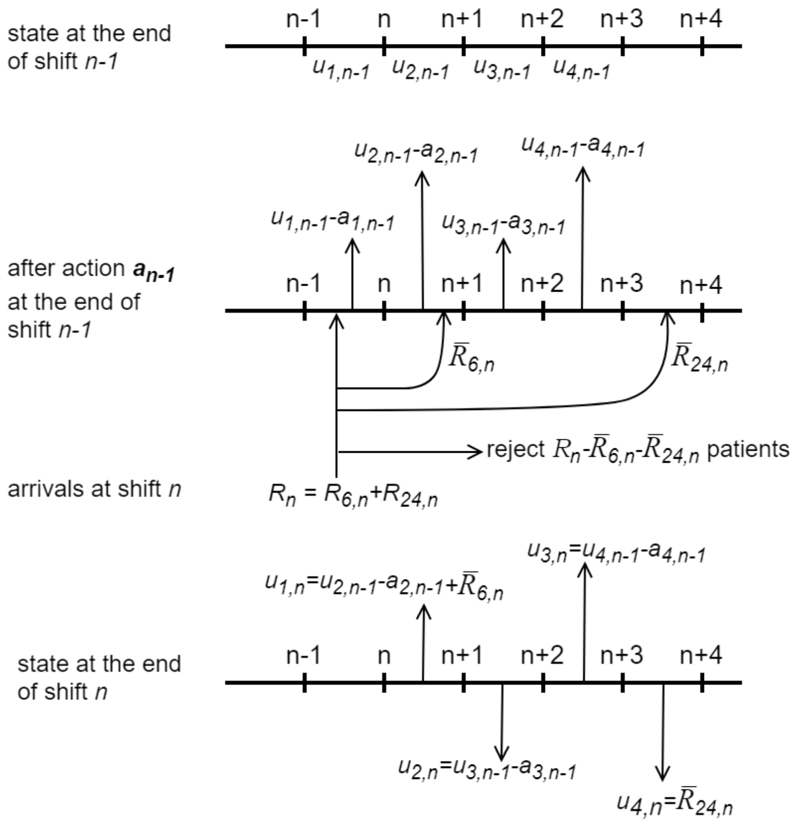


Figure 4.2: State evolution from shift $n-1$ to shift n

we admit $\bar{R}_{24,n}$ 24-hour arrivals, where $\bar{R}_{24,n} = \min\{R_{24,n}, 4s - (u_{1,n} + u_{2,n} + u_{3,n})\}$. Hence, $u_{4,n} = \min\{R_{24,n}, 4s - (u_{1,n} + u_{2,n} + u_{3,n})\}$.

To formulate the transition probability further, the definition of the exponential distribution is given in Definition 4.3.3. The definition of the Erlang distribution is shown in Definition 4.3.4.

Definition 4.3.3. *[21] A continuous random variable X is said to have an exponential distribution with parameter λ, $\lambda > 0$ if its probability density function is given by*

$$f(x) = \lambda e^{-\lambda x},\ x \geq 0 \tag{4.4}$$

or, equivalently, if its cumulative distribution function is given by

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = 1 - e^{-\lambda x},\ x \geq 0. \tag{4.5}$$

Definition 4.3.4. *A random variable X has an Erlang-k ($k = 1, 2, \ldots$) distribution with mean k/λ, if X is the sum of k independent random variables $X_1, \ldots, X_k$ having an exponential distribution with mean $1/\lambda$ and the probability density function is given* *by*

$$f_{Erl}(x;k,\lambda) = \lambda \frac{(\lambda\, x)^{k-1}}{(k-1)!} e^{-\lambda\, x},\ x \geq 0. \tag{4.6}$$

The cumulative distribution function is given by

$$F_{Erl}(x;k,\lambda) = 1 - \sum_{j=0}^{k-1} \frac{(\lambda\, x)^{j}}{j!} e^{-\lambda\, x},\ x \geq 0. \tag{4.7}$$

The parameter λ is called the scale parameter, k is the shape parameter.

For the transition probability, we use the notation below:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{u_n} \,|\, \boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}) =& \mathbb{P}(U_{1,n} = u_{1,n},\, U_{2,n} = u_{2,n},\, U_{3,n} = u_{3,n}, U_{4,n} = u_{4,n} | \\
& U_{1,n-1} = u_{1,n-1},\, U_{2,n-1} = u_{2,n-1},\, U_{3,n-1} = u_{3,n-1},\, U_{4,n-1} = u_{4,n-1}, \\
& \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1})) \\
=& \mathbb{P}(u_{1,n} = \min\{u_{2,n-1} - a_{2,n-1} + R_{6,n}, s\},\, u_{2,n} = u_{3,n-1} - a_{3,n-1}, \\
& u_{3,n} = u_{4,n-1} - a_{4,n-1}, u_{4,n} = \min\{R_{24,n}, 4s - (u_{1,n} + u_{2,n} + u_{3,n})\} | \\
& R_{6,n} \geq 0,\, R_{24,n} \geq 0).
\end{aligned} \tag{4.8}$$

Employing Assumption 4.3.2 and state evolution in Figure 4.2, we build the transition probabilities based on the patient arrivals.

If $u_{2,n-1} - a_{2,n-1} + R_{6,n} < s$, then at the end of shift n the first element of the state is $u_{1,n} = u_{2,n-1} - a_{2,n-1} + R_{6,n} < s$. After that, if $u_{1,n} + u_{2,n} + u_{3,n} + R_{24,n} < 4s$ then $u_{4,n} = R_{24,n}$. Hence, the state at the end of shift n is $u_{1,n} < s,\ u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} < 4s$, which transition probability is as follows:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{u_n} \,|\, \boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}) =& \mathbb{P}(U_{1,n} = u_{1,n},\, U_{2,n} = u_{2,n},\, U_{3,n} = u_{3,n}, U_{4,n} = u_{4,n} | \\
& U_{1,n-1} = u_{1,n-1},\, U_{2,n-1} = u_{2,n-1},\, U_{3,n-1} = u_{3,n-1},\, U_{4,n-1} = u_{4,n-1}, \\
& \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1})) \\
=& \mathbb{P}(u_{1,n} = u_{2,n-1} - a_{2,n-1} + R_{6,n},\, u_{2,n} = u_{3,n-1} - a_{3,n-1}, \\
& u_{3,n} = u_{4,n-1} - a_{4,n-1},\, u_{4,n} = R_{24,n}) \\
=& \mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1}\ ,\ R_{24,n} = u_{4,n}) \\
=& \mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1})\ \times\ \mathbb{P}(R_{24,n} = u_{4,n}) \\
=& e^{-\lambda_6}\, e^{-\lambda_{24}}\, \frac{\lambda_6^{u_{1,n} - u_{2,n-1} + a_{2,n-1}}}{(u_{1,n} - u_{2,n-1} + a_{2,n-1})!} \cdot \frac{\lambda_{24}^{u_{4,n}}}{u_{4,n}!}, \\
=& Poi(u_{1,n} - u_{2,n-1} + a_{2,n-1}, \lambda_6)\, .\, Poi(u_{4,n}, \lambda_{24})
\end{aligned} \tag{4.9}$$

$$u_{1,n} < s\ ;\ u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} < 4s.$$

If $u_{2,n-1} - a_{2,n-1} + R_{6,n} < s$, then at the end of shift n the first element of the state is $u_{1,n} = u_{2,n-1} - a_{2,n-1} + R_{6,n} < s$. Next, if $u_{1,n} + u_{2,n} + u_{3,n} + R_{24,n} \geq 4s$, then at the end of shift n, the state is given by $u_{1,n} < s$, $u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} = 4s$, which is equivalent to $u_{1,n} < s, u_{4,n} = 4s - (u_{1,n} + u_{2,n} + u_{3,n})$. Thus, for this case we have the following transition probability:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{u_n} \,|\, \boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}) =& \mathbb{P}(U_{1,n} = u_{1,n},\, U_{2,n} = u_{2,n},\, U_{3,n} = u_{3,n},\, U_{4,n} = u_{4,n}| \\
& U_{1,n-1} = u_{1,n-1},\, U_{2,n-1} = u_{2,n-1},\, U_{3,n-1} = u_{3,n-1},\, U_{4,n-1} = u_{4,n-1}, \\
& \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1})) \\
=& \mathbb{P}(u_{1,n} = u_{2,n-1} - a_{2,n-1} + R_{6,n},\, u_{2,n} = u_{3,n-1} - a_{3,n-1}, \\
& u_{3,n} = u_{4,n-1} - a_{4,n-1},\, u_{4,n} = 4s - (u_{1,n} + u_{2,n} + u_{3,n})) \\
=& \mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1}\, ,\, u_{2,n} = u_{3,n-1} - a_{3,n-1}, \\
& \qquad u_{3,n} = u_{4,n-1} - a_{4,n-1}\, ,\, R_{24,n} \geq 4s - (u_{1,n} + u_{2,n} + u_{3,n})) \\
=& \mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1}\, , \\
& \qquad\qquad R_{24,n} \geq 4s - (u_{1,n} + u_{2,n} + u_{3,n})) \\
=& \mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1}) \\
& \qquad \times\ \mathbb{P}(R_{24,n} \geq 4s - (u_{1,n} + u_{2,n} + u_{3,n})) \\
=& e^{-\lambda_6} \frac{\lambda_6^{u_{1,n} - u_{2,n-1} + a_{2,n-1}}}{(u_{1,n} - u_{2,n-1} + a_{2,n-1})!} \cdot \sum_{r_{24} = 4s - (u_{1,n} + u_{2,n} + u_{3,n})}^{\infty} e^{-\lambda_{24}} \frac{\lambda_{24}^{r_{24}}}{r_{24}!} \\
=& e^{-\lambda_6} \frac{\lambda_6^{u_{1,n} - u_{2,n-1} + a_{2,n-1}}}{(u_{1,n} - u_{2,n-1} + a_{2,n-1})!} \cdot \left(1 - \sum_{r_{24} = 0}^{4s - (u_{1,n} + u_{2,n} + u_{3,n}) - 1} e^{-\lambda_{24}} \frac{\lambda_{24}^{r_{24}}}{r_{24}!} \right) \\
=& Poi(u_{1,n} - u_{2,n-1} + a_{2,n-1}, \lambda_6) \,.\, F_{Erl}\left(4s - (u_{1,n} + u_{2,n} + u_{3,n}),\, \lambda_{24}\right),
\end{aligned} \tag{4.10}$$

$$u_{1,n} < s\,;\, u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} = 4s,$$

where $F_{Erl}(k, \lambda)$ is the cumulative distribution function of the Erlang-k distribution with scale parameter λ.

If $u_{2,n-1} - a_{2,n-1} + R_{6,n} \geq s$, then at the end of shift n the first element of the state is $u_{1,n} = s$. Next, if $u_{1,n} + u_{2,n} + u_{3,n} + R_{24,n} < 4s$, then at the end of shift n the last element of the state is $u_{4,n} = R_{24,n}$. Hence, the state at the end of shift n is $u_{1,n} = s$, $u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} < 4s$ and the transition probability is given by:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{u_n} \,|\, \boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}) =& \mathbb{P}(U_{1,n} = u_{1,n},\, U_{2,n} = u_{2,n},\, U_{3,n} = u_{3,n},\, U_{4,n} = u_{4,n}| \\
& U_{1,n-1} = u_{1,n-1},\, U_{2,n-1} = u_{2,n-1},\, U_{3,n-1} = u_{3,n-1},\, U_{4,n-1} = u_{4,n-1}, \\
& \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1}))
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}(u_{1,n} = s,\, u_{2,n} = u_{3,n-1} - a_{3,n-1},\\
&\quad u_{3,n} = u_{4,n-1} - a_{4,n-1},\, u_{4,n} = R_{24,n})\\
&= \mathbb{P}(s \le u_{2,n-1} - a_{2,n-1} + R_{6,n},\, u_{2,n} = u_{3,n-1} - a_{3,n-1},\\
&\quad u_{3,n} = u_{4,n-1} - a_{4,n-1},\, u_{4,n} = R_{24,n})\\
&= \mathbb{P}(R_{6,n} \ge s - u_{2,n-1} + a_{2,n-1}\,,\, R_{24,n} = u_{4,n})\\
&= \mathbb{P}(R_{6,n} \ge s - u_{2,n-1} + a_{2,n-1}) \;\times\; \mathbb{P}(R_{24,n} = u_{4,n})\\
&= e^{-\lambda_{24}} \sum_{r_6 = s - u_{2,n-1} + a_{2,n-1}}^{\infty} e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!} \cdot \frac{\lambda_{24}^{u_{4,n}}}{u_{4,n}!}\\
&= e^{-\lambda_{24}} \frac{\lambda_{24}^{u_{4,n}}}{u_{4,n}!} \left(1 - \sum_{r_6=0}^{s - u_{2,n-1} + a_{2,n-1} - 1} e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!}\right)\\
&= Poi(u_{4,n}, \lambda_{24})\,.\, F_{Erl}\left(s - u_{2,n-1} + a_{2,n-1},\, \lambda_6\right) \qquad (4.11)
\end{aligned}
$$

$$u_{1,n} = s\ ;\ u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} < 4s.$$

If $u_{2,n-1} - a_{2,n-1} + R_{6,n} \ge s$, then at the end of shift n the first element of the state is $u_{1,n} = s$. Next, if $u_{1,n} + u_{2,n} + u_{3,n} + R_{24,n} \ge 4s$, then at the end of shift n the state is given by $u_{1,n} = s$, $u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} = 4s$, which is equivalent to $u_{1,n} = s$, $u_{4,n} = 3s - (u_{2,n} + u_{3,n})$ and the transition probability is given by:

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{u_n} \,|\, \boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}) &= \mathbb{P}(U_{1,n} = u_{1,n},\, U_{2,n} = u_{2,n},\, U_{3,n} = u_{3,n},\, U_{4,n} = u_{4,n}|\\
&\quad U_{1,n-1} = u_{1,n-1},\, U_{2,n-1} = u_{2,n-1},\, U_{3,n-1} = u_{3,n-1},\, U_{4,n-1} = u_{4,n-1},\\
&\quad \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1}))\\
&= \mathbb{P}(u_{1,n} = s,\, u_{2,n} = u_{3,n-1} - a_{3,n-1},\\
&\quad u_{3,n} = u_{4,n-1} - a_{4,n-1},\, u_{4,n} = 3s - (u_{2,n} + u_{3,n}))\\
&= \mathbb{P}(s \le u_{2,n-1} - a_{2,n-1} + R_{6,n},\, u_{2,n} = u_{3,n-1} - a_{3,n-1},\\
&\quad u_{3,n} = u_{4,n-1} - a_{4,n-1},\, 3s - (u_{2,n} + u_{3,n}) \le R_{24,n})\\
&= \mathbb{P}(R_{6,n} \ge s - u_{2,n-1} + a_{2,n-1}\,,\, R_{24,n} \ge 3s - u_{2,n} - u_{3,n})\\
&= \mathbb{P}(R_{6,n} \ge s - u_{2,n-1} + a_{2,n-1}) \;\times\; \mathbb{P}(R_{24,n} \ge 3s - u_{2,n} - u_{3,n})\\
&= e^{-\lambda_6}\, e^{-\lambda_{24}} \sum_{r_6 = s - u_{2,n-1} + a_{2,n-1}}^{\infty} \frac{\lambda_6^{r_6}}{r_6!} \cdot \sum_{r_{24} = 3s - u_{2,n} - u_{3,n}}^{\infty} \frac{\lambda_{24}^{r_{24}}}{r_{24}!}\\
&= \left(1 - \sum_{r_6=0}^{s - u_{2,n-1} + a_{2,n-1} - 1} e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!}\right) \cdot \left(1 - \sum_{r_{24}=0}^{3s - u_{2,n} - u_{3,n} - 1} e^{-\lambda_{24}} \frac{\lambda_{24}^{r_{24}}}{r_{24}!}\right)\\
&= F_{Erl}\left(s - u_{2,n-1} + a_{2,n-1},\, \lambda_6\right) \,.\, F_{Erl}\left(3s - u_{2,n} - u_{3,n},\, \lambda_{24}\right), \qquad (4.12)
\end{aligned}
$$

$$u_{1,n} = s\ ;\ u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} = 4s.$$

Direct cost

The first cost incurred because of the rejections of the 6-hour patient arrivals from entering the system. From Assumption 4.3.2 we know that if $u_{1,n} \leq s$, all 6-hour patients can be treated in the dedicated capacity. If $u_{2,n-1} - a_{2,n-1} + R_{6,n} > s$ we need to reject $u_{2,n-1} - a_{2,n-1} + R_{6,n} - s$ patients. Note that at shift n, $u_{1,n}$ depends on the state and action at shift $n-1$ as $u_{1,n} = \min\{u_{2,n-1} - a_{2,n-1} + R_{6,n}, s\}$, where $R_{6,n}$ is the 6-hour patient arrivals at shift n. Denoting the number of 6-hour surgeries being rejected at shift n by $N_{e,n}$, the formula to compute its expectation, $\mathbb{E}[N_{e,n}|\boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}]$, is:

$$\begin{aligned}
\mathbb{E}[N_{e,n}|\boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}] &= \mathbb{E}[u_{2,n-1} - a_{2,n-1} + R_{6,n} - s]^+ \\
&= \sum_{r_6=0}^{\infty} (u_{2,n-1} - a_{2,n-1} + r_6 - s)^+ \, \mathbb{P}(R_{6,n} = r_6) \\
&= \sum_{r_6=s-u_{2,n-1}+a_{2,n-1}+1}^{\infty} (u_{2,n-1} - a_{2,n-1} + r_6 - s) \, \mathbb{P}(R_{6,n} = r_6) \\
&= \sum_{r_6=s-u_{2,n-1}+a_{2,n-1}+1}^{\infty} (u_{2,n-1} - a_{2,n-1} - s) \, \mathbb{P}(R_{6,n} = r_6) \\
&\qquad + \sum_{r_6=s-u_{2,n-1}+a_{2,n-1}+1}^{\infty} r_6 \, \mathbb{P}(R_{6,n} = r_6) \\
&= (u_{2,n-1} - a_{2,n-1} - s) \left(1 - \sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} \mathbb{P}(R_{6,n} = r_6)\right) \\
&\qquad + \sum_{r_6=0}^{\infty} r_6 \, \mathbb{P}(R_{6,n} = r_6) - \sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} r_6 \, \mathbb{P}(R_{6,n} = r_6) \\
&= (u_{2,n-1} - a_{2,n-1} - s) \left(1 - \sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!}\right) + \mathbb{E}[R_{6,n}] \\
&\qquad - \sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} r_6 \, e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!} \\
&= (u_{2,n-1} - a_{2,n-1} - s) \left(1 - \sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!}\right) + \lambda_6 \\
&\qquad - \sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} r_6 \, e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!}.
\end{aligned} \tag{4.13}$$

Let $N = u_{2,n-1} - a_{2,n-1} - s$. We have:

$$\sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} r_6 \, e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!} = \sum_{n=0}^{N} n \, e^{-\lambda_6} \frac{\lambda_6^{n}}{n!}, \tag{4.14}$$

and is elaborated as follows:

$$\begin{aligned}\sum_{n=0}^{N} n\ e^{-\lambda_6}\,\frac{\lambda_6^n}{n!} =&\sum_{n=1}^{N} n\ e^{-\lambda_6}\,\frac{\lambda_6^n}{n!}\\ =&\sum_{n=1}^{N}\ e^{-\lambda_6}\,\frac{\lambda_6^n}{(n-1)!}\\ =&\lambda_6\sum_{n=1}^{N}\ e^{-\lambda_6}\,\frac{\lambda_6^{n-1}}{(n-1)!}\\ =&\lambda_6\sum_{n=1}^{N}\ e^{-\lambda_6}\,\frac{\lambda_6^{n-1}}{(n-1)!}\\ =&\lambda_6\sum_{m=0}^{N-1}\ e^{-\lambda_6}\,\frac{\lambda_6^{m}}{m!}\\ =&\lambda_6(1-F_{Erl}(N,\lambda_6)).\end{aligned} \tag{4.15}$$

Substituting Equation (4.15) to Equation (4.13), we obtain:

$$\begin{aligned}\mathbb{E}[N_{e,n}|\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}] &= (u_{2,n-1}-a_{2,n-1}-s)\left(1-\sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}} e^{-\lambda_6}\,\frac{\lambda_6^{r_6}}{r_6!}\right)+\lambda_6\\ &\quad -\lambda_6(1-F_{Erl}(s-u_{2,n-1}+a_{2,n-1},\lambda_6))\\ &=(u_{2,n-1}-a_{2,n-1}-s)\,F_{Erl}(s-u_{2,n-1}+a_{2,n-1}+1,\lambda_6)\\ &\quad +\lambda_6\,F_{Erl}(s-u_{2,n-1}+a_{2,n-1},\lambda_6),\forall n\in\mathbb{N}.\end{aligned} \tag{4.16}$$

Another cost is incurred from rejecting the 24-hour surgeries because of the boundary of the state. Denoting the number of 24-hour surgeries being rejected at shift n by $\hat{N}_{e,n}$, it is formulated as follows:

$$\hat{N}_{e,n} = (u_{1,n}+u_{2,n}+u_{3,n}+R_{24,n}-4s)^+,\forall n\in\mathbb{N}. \tag{4.17}$$

Recall that the state at shift n depends on the state $\boldsymbol{u_{n-1}}$ and action $\boldsymbol{a_{n-1}}$. Hence, $\mathbb{E}[\hat{N}_{e,n}|\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}]$ is formulated as follows.

$$\begin{aligned}\mathbb{E}[\hat{N}_{e,n}|\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}] &= \mathbb{E}[(u_{1,n}+u_{2,n}+u_{3,n}+R_{24,n}-4s)^+|\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}]\\ &=\mathbb{E}[(u_{2,n-1}-a_{2,n-1}+R_{6,n}+u_{2,n}+u_{3,n}+R_{24,n}-4s)^+]\\ &=\sum_{r_6=0}^{\infty}\sum_{r_{24}=0}^{\infty}(u_{2,n-1}-a_{2,n-1}+r_6+u_{2,n}+u_{3,n}+r_{24}-4s)^+\\ &\qquad \mathbb{P}(R_{6,n}=r_6)\,\mathbb{P}(R_{24,n}=r_{24})\end{aligned}$$

$$
\begin{aligned}
&= \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} \sum_{\substack{r_{24}=4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})+1}}^{\infty} (u_{2,n-1} - a_{2,n-1} + r_6 + u_{2,n} + u_{3,n} \\
&\qquad + r_{24} - 4s)\, \mathbb{P}(R_{6,n} = r_6)\, \mathbb{P}(R_{24,n} = r_{24}) \\
&\qquad + \sum_{r_6=s-(u_{2,n-1}-a_{2,n-1})}^{\infty} \sum_{\substack{r_{24}=3s-(u_{2,n}\\ +u_{3,n})+1}}^{\infty} (s + u_{2,n} + u_{3,n} \\
&\qquad\qquad + r_{24} - 4s)\, \mathbb{P}(R_{6,n} = r_6)\, \mathbb{P}(R_{24,n} = r_{24}) \\
&= (u_{2,n-1} - a_{2,n-1} + u_{2,n} + u_{3,n} - 4s) \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} \mathbb{P}(R_{6,n} = r_6) \\
&\qquad \sum_{\substack{r_{24}=4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})+1}}^{\infty} \mathbb{P}(R_{24,n} = r_{24}) \\
&+ \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} \sum_{\substack{r_{24}=4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})+1}}^{\infty} r_6\, \mathbb{P}(R_{6,n} = r_6)\mathbb{P}(R_{24,n} = r_{24}) \\
&\quad + \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} \sum_{\substack{r_{24}=4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})+1}}^{\infty} r_{24}\, \mathbb{P}(R_{6,n} = r_6)\mathbb{P}(R_{24,n} = r_{24}) \\
&\qquad + (s + u_{2,n} + u_{3,n} - 4s) \\
&\sum_{r_6=s-(u_{2,n-1}-a_{2,n-1})}^{\infty} \sum_{\substack{r_{24}=3s-(u_{2,n}\\ +u_{3,n})+1}}^{\infty} \mathbb{P}(R_{6,n} = r_6)\mathbb{P}(R_{24,n} = r_{24}) \\
&+ \sum_{r_6=s-(u_{2,n-1}-a_{2,n-1})}^{\infty} \mathbb{P}(R_{6,n} = r_6) \sum_{\substack{r_{24}=3s-(u_{2,n}\\ +u_{3,n})+1}}^{\infty} r_{24}\, \mathbb{P}(R_{24,n} = r_{24}) \\
&= (u_{2,n-1} - a_{2,n-1} + u_{2,n} + u_{3,n} - 4s) \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} \mathbb{P}(R_{6,n} = r_6) \\
&\qquad \left(1 - \sum_{r_{24}=0}^{\substack{4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})}} \mathbb{P}(R_{24,n} = r_{24}) \right)
\end{aligned}
$$

$$+ \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} r_6\,\mathbb{P}(R_{6,n}=r_6)\left(1-\sum_{r_{24}=0}^{\substack{4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})}}\mathbb{P}(R_{24,n}=r_{24})\right)$$

$$+ \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} \mathbb{P}(R_{6,n}=r_6)\left(\sum_{r_{24}=0}^{\infty} r_{24}\,\mathbb{P}(R_{24,n}=r_{24})-\sum_{r_{24}=0}^{\substack{4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})+1}} r_{24}\,\mathbb{P}(R_{24,n}=r_{24})\right)$$

$$+ (u_{2,n}+u_{3,n}-3s)\left(1-\sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}}\mathbb{P}(R_{6,n}=r_6)\right)\left(1-\sum_{r_{24}=0}^{\substack{3s-(u_{2,n}\\ +u_{3,n})}}\mathbb{P}(R_{24,n}=r_{24}\right)$$

$$+ \left(1-\sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}}\mathbb{P}(R_{6,n}=r_6)\right)\left(\sum_{r_{24}=0}^{\infty} r_{24}\,\mathbb{P}(R_{24,n}=r_{24})-\sum_{r_{24}=0}^{\substack{3s-(u_{2,n}\\ +u_{3,n})}} r_{24}\,\mathbb{P}(R_{24,n}=r_{24})\right)$$

$$= (u_{2,n-1}-a_{2,n-1}+u_{2,n}+u_{3,n}-4s)\sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}}\mathbb{P}(R_{6,n}=r_6)$$

$$\left(1-\sum_{r_{24}=0}^{\substack{4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})}}\mathbb{P}(R_{24,n}=r_{24})\right)$$

$$+ \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} r_6\,\mathbb{P}(R_{6,n}=r_6)\left(1-\sum_{r_{24}=0}^{\substack{4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})}}\mathbb{P}(R_{24,n}=r_{24})\right)$$

$$+ \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} \mathbb{P}(R_{6,n}=r_6)\left(\lambda_{24}-\sum_{r_{24}=0}^{\substack{4s-(u_{2,n-1}\\ -a_{2,n-1}+r_6+u_{2,n}\\ +u_{3,n})+1}} r_{24}\,\mathbb{P}(R_{24,n}=r_{24})\right)$$

$$+ (u_{2,n}+u_{3,n}-3s)\left(1-\sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}}\mathbb{P}(R_{6,n}=r_6)\right)\left(1-\sum_{r_{24}=0}^{\substack{3s-(u_{2,n}\\ +u_{3,n})}}\mathbb{P}(R_{24,n}=r_{24}\right)$$

$$+ \left(1-\sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}}\mathbb{P}(R_{6,n}=r_6)\right)\left(\lambda_{24}-\sum_{r_{24}=0}^{\substack{3s-(u_{2,n}\\ +u_{3,n})}} r_{24}\,\mathbb{P}(R_{24,n}=r_{24})\right)$$

$$
\begin{aligned}
&= (u_{2,n-1} - a_{2,n-1} + u_{2,n} + u_{3,n} - 4s) \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} Poi(r_6, \lambda_6) \\
&\qquad F_{Erl}(4s - (u_{2,n-1} - a_{2,n-1} + r_6 + u_{2,n} + u_{3,n}) + 1, \lambda_{24}) \\
&\quad + \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} r_6 \, Poi(r_6, \lambda_6) \, F_{Erl}(4s - (u_{2,n-1} - a_{2,n-1} + r_6 + u_{2,n} + u_{3,n}) + 1, \lambda_{24}) \\
&+ \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} Poi(r_6, \lambda_6)(\lambda_{24} - \lambda_{24}(1 - F_{Erl}(4s - (u_{2,n-1} - a_{2,n-1} + r_6 + u_{2,n} + u_{3,n}) + 1, \\
&\quad \lambda_{24}))) + (u_{2,n} + u_{3,n} - 3s) \, F_{Erl}(s - (u_{2,n-1} - a_{2,n-1}), \lambda_6) \, F_{Erl}(3s - (u_{2,n} + u_{3,n}) + 1, \\
&\quad \lambda_{24}) + F_{Erl}(s - (u_{2,n-1} - a_{2,n-1}), \lambda_6)(\lambda_{24} - \lambda_{24}(1 - F_{Erl}(3s - (u_{2,n} + u_{3,n}), \lambda_{24}))) \\
&= (u_{2,n-1} - a_{2,n-1} + u_{2,n} + u_{3,n} - 4s) \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} Poi(r_6, \lambda_6) \\
&\qquad F_{Erl}(4s - (u_{2,n-1} - a_{2,n-1} + r_6 + u_{2,n} + u_{3,n}) + 1, \lambda_{24}) \\
&+ \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} r_6 \, Poi(r_6, \lambda_6) \, F_{Erl}(4s - (u_{2,n-1} - a_{2,n-1} + r_6 + u_{2,n} + u_{3,n}) + 1, \lambda_{24}) \\
&\quad + \sum_{r_6=0}^{\substack{s-(u_{2,n-1}\\ -a_{2,n-1})-1}} Poi(r_6, \lambda_6) \, \lambda_{24} \, F_{Erl}(4s - (u_{2,n-1} - a_{2,n-1} + r_6 + u_{2,n} + u_{3,n}) + 1, \lambda_{24}) \\
&\quad + (u_{2,n} + u_{3,n} - 3s) \, F_{Erl}(s - (u_{2,n-1} - a_{2,n-1}), \lambda_6) \, F_{Erl}(3s - (u_{2,n} + u_{3,n}) + 1, \lambda_{24}) \\
&\qquad + F_{Erl}(s - (u_{2,n-1} - a_{2,n-1}), \lambda_6) \, \lambda_{24} \, F_{Erl}(3s - (u_{2,n} + u_{3,n}), \lambda_{24}).
\end{aligned} \tag{4.18}
$$

Given the state $\boldsymbol{u_n}$ and action $\boldsymbol{a_n}$ the number of unused capacity $N_{u,n}$ can be formulated as follows:

$$
N_{u,n} = \left(s - \sum_{t=1}^{4} a_{t,n} \right)^{+} \mathbb{1} \left(\sum_{t=1}^{4} u_{t,n} > 0 \right), \forall n \in \mathbb{N}. \tag{4.19}
$$

Let c_e, $\hat{c}_e$, and c_u denote the cost of one rejected 6-hour patient, one rejected 24-hour patient, and one unused dedicated capacity, respectively. Henceforth, by using Equations (4.16), (4.19) and (4.18) we obtain the expected total costs in shift n is:

$$
\mathbb{E}[c_n] = c_e \, \mathbb{E}[N_{e,n}] + \hat{c}_e \, \mathbb{E}[\hat{N}_{e,n}] + c_u \, \mathbb{E}[N_{u,n}], \;\; \forall n \in \mathbb{N}. \tag{4.20}
$$

Optimality equation

We consider the costs incurred today to be more important than those that are incurred tomorrow. Therefore we use discount factor $\alpha, \alpha \in [0, 1)$ to recalculate the future costs to the cost level today. The goal is to minimize the rejected patients and utilize the dedicated capacity optimally. This can be done by minimizing the expected discounted cost over an infinite horizon, because the costs are incurred from the unused dedicated capacity and rejecting patients from entering the system. Therefore, the optimality equation is as follows:

$$\boldsymbol{V}(\boldsymbol{u}) = \min_{\boldsymbol{a}\in A}\left\{c(\boldsymbol{u}, \boldsymbol{a_0}) + \alpha\sum_{\boldsymbol{u'}}\mathbb{P}(\boldsymbol{u'}|\boldsymbol{u}, \boldsymbol{a})\,\boldsymbol{V}(\boldsymbol{u'})\right\}. \tag{4.21}$$

The optimal policy A^* consists of the values of a that solve the optimality equation in each state. We use the policy iteration algorithm to find the optimal policy A^*.

Algorithm 1 Policy iteration algorithm

```
$i \leftarrow 0$
$A_0 \leftarrow \bar{A}$                                        ▷ Choose an arbitrary policy $\bar{A} \in A$
while $A_{i+1} \neq A_i$ do
    for each state $\boldsymbol{u_n} \in \mathcal{S}$ do
        Policy Evaluation   : $V_i(\boldsymbol{u_n}) = c(\boldsymbol{u_n}, A_i) + \alpha\sum_{\boldsymbol{u'_n}} P(\boldsymbol{u'_n}|\boldsymbol{u_n}, A_i)\, V(\boldsymbol{u'_n})$
        Policy improvement :
        $A_{i+1} \in \arg\min_{A}\{c(\boldsymbol{u_n}, A) + \alpha\sum_{\boldsymbol{u'_n}} \mathbb{P}(\boldsymbol{u'_n}|\boldsymbol{u_n}, A)\, V_i(u'_n)\}$
    end for
    $i \leftarrow i + 1$
end while
return $A^* = A_i$                                          ▷ The optimal policy $A^*$
```

4.3.1 Enable deferring the 24-hour patients to another resource (Model 1b)

In the model explained before, we determine the number of surgeries that are performed in every shift in the dedicated capacity. To extend that model, in this part we allow some of the 24-hour patients are deferred to other resources, such as the elective operating rooms. Deferring 24-hour patients to another resource incurs a certain cost. Deferring too many 24-hour patients to other resources incurs a big cost and might leave the dedicated capacity unused, while deferring too few of them causes more 6-hour patients to be rejected from the dedicated capacity.

In this part the state of the system is described by Equation (4.2). The decisions now include deferring the 24-hour patients to another resource, denoted by $b_{4,n}$. Note that the total 24-hour surgeries that are performed and deferred to other resource should not exceed the existing 24-hour patients in the system, formulated by $a_{4,n} + b_{4,n} \leq u_{4,n}$. With this modification, the decisions are denoted as follows:

$$A_{\boldsymbol{u},n} = \{\boldsymbol{a_n} = (a_{1,n}, a_{2,n}, a_{3,n}, a_{4,n}, b_{4,n}) = (\text{do } a_{t,n} \text{ surgery out of } u_{t,n},\ t = 1, 2, 3,$$
$$\text{deferring } b_{4,n} \text{ 24-hour patients}) \,|\, a_{t,n} \leq u_{t,n},\ t = 1, 2, 3\,;\, a_{4,n} + b_{4,n} \leq u_{4,n},$$
$$\sum_{t=1}^{4} a_{t,n} \leq s,\ \forall n \in \mathbb{N}\},$$
$$A = \bigcup_{\boldsymbol{u} \in \mathcal{S}, n \in \mathbb{N}} A_{\boldsymbol{u},n}.$$

The evolution of the states is given in the diagram in Figure 4.3.

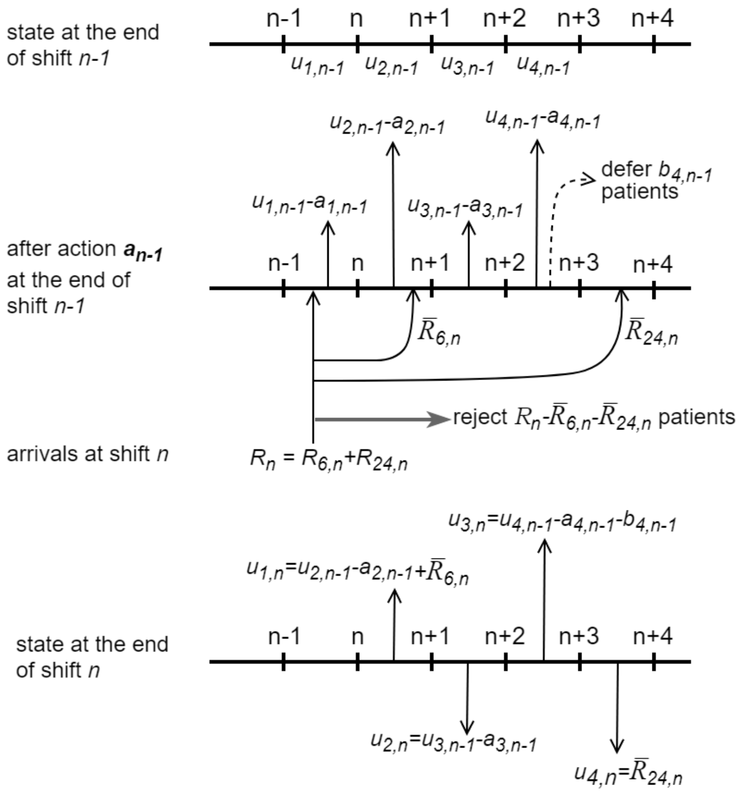


Figure 4.3: State evolution from the beginning of shift $n-1$ to the end of shift n while allowing 24-hour patients to be deferred

Considering Assumption 4.3.2 and the state evolution in Figure 4.3, we construct the transition probability. The following notation is used :

$$\begin{aligned}
\mathbb{P}(\boldsymbol{u_n}\,|\,\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}) =&\mathbb{P}(U_{1,n}=u_{1,n},\, U_{2,n}=u_{2,n},\, U_{3,n}=u_{3,n}, U_{4,n}=u_{4,n}|\\
&U_{1,n-1}=u_{1,n-1},\, U_{2,n-1}=u_{2,n-1},\, U_{3,n-1}=u_{3,n-1},\, U_{4,n-1}=u_{4,n-1},\\
&\boldsymbol{a_{n-1}}=(a_{1,n-1},a_{2,n-1},a_{3,n-1},a_{4,n-1},b_{4,n-1}))\\
=&\mathbb{P}(u_{1,n}=\min\{u_{2,n-1}-a_{2,n-1}+R_{6,n},s\},\, u_{2,n}=u_{3,n-1}-a_{3,n-1},\\
&u_{3,n}=u_{4,n-1}-a_{4,n-1}-b_{4,n-1},u_{4,n}=\min\{R_{24,n},4s-(u_{1,n}+\\
&u_{2,n}+u_{3,n})\}|R_{6,n}\geq 0,\, R_{24,n}\geq 0). \qquad (4.22)
\end{aligned}$$

Based on the arrivals, if $u_{2,n-1}-a_{2,n-1}+R_{6,n}<s$, then at the end of shift n the first element of the state is $u_{1,n}=u_{2,n-1}-a_{2,n-1}+R_{6,n}<s$. After that, if $u_{1,n}+u_{2,n}+u_{3,n}+R_{24,n}<4s$ then $u_{4,n}=R_{24,n}$. Hence, the state at the end of shift n is $u_{1,n}<s,\, u_{1,n}+u_{2,n}+u_{3,n}+u_{4,n}<4s$, which transition probability is as follows:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{u_n}\,|\,\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}) =&\mathbb{P}(U_{1,n}=u_{1,n},\, U_{2,n}=u_{2,n},\, U_{3,n}=u_{3,n}, U_{4,n}=u_{4,n}|\\
&U_{1,n-1}=u_{1,n-1},\, U_{2,n-1}=u_{2,n-1},\, U_{3,n-1}=u_{3,n-1},\, U_{4,n-1}=u_{4,n-1},\\
&\boldsymbol{a_{n-1}}=(a_{1,n-1},a_{2,n-1},a_{3,n-1},a_{4,n-1},b_{4,n-1}))\\
=&\mathbb{P}(u_{1,n}=u_{2,n-1}-a_{2,n-1}+R_{6,n},\, u_{2,n}=u_{3,n-1}-a_{3,n-1},\\
&u_{3,n}=u_{4,n-1}-a_{4,n-1}-b_{4,n-1},\, u_{4,n}=R_{24,n})\\
=&\mathbb{P}(R_{6,n}=u_{1,n}-u_{2,n-1}+a_{2,n-1}\,,\, R_{24,n}=u_{4,n})\\
=&\mathbb{P}(R_{6,n}=u_{1,n}-u_{2,n-1}+a_{2,n-1})\ \times\ \mathbb{P}(R_{24,n}=u_{4,n})\\
=&e^{-\lambda_6}\, e^{-\lambda_{24}}\ \frac{\lambda_6^{u_{1,n}-u_{2,n-1}+a_{2,n-1}}}{(u_{1,n}-u_{2,n-1}+a_{2,n-1})!}\cdot\frac{\lambda_{24}^{u_{4,n}}}{u_{4,n}!},\\
=&Poi(u_{1,n}-u_{2,n-1}+a_{2,n-1},\lambda_6)\,.\,Poi(u_{4,n},\lambda_{24}) \qquad (4.23)\\
&u_{1,n}<s\ ;\ u_{1,n}+u_{2,n}+u_{3,n}+u_{4,n}<4s.
\end{aligned}$$

If $u_{2,n-1}-a_{2,n-1}+R_{6,n}<s$, then at the end of shift n the first element of the state is $u_{1,n}=u_{2,n-1}-a_{2,n-1}+R_{6,n}<s$. Next, if $u_{1,n}+u_{2,n}+u_{3,n}+R_{24,n}\geq 4s$, then at the end of shift n, the state is given by $u_{1,n}<s,\, u_{1,n}+u_{2,n}+u_{3,n}+u_{4,n}=4s$, which is equivalent to $u_{1,n}<s, u_{4,n}=4s-(u_{1,n}+u_{2,n}+u_{3,n})$. Thus, for this case we have the following transition probability:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{u_n}\,|\,\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}) =&\mathbb{P}(U_{1,n}=u_{1,n},\, U_{2,n}=u_{2,n},\, U_{3,n}=u_{3,n},\, U_{4,n}=u_{4,n}|\\
&U_{1,n-1}=u_{1,n-1},\, U_{2,n-1}=u_{2,n-1},\, U_{3,n-1}=u_{3,n-1},\, U_{4,n-1}=u_{4,n-1},\\
&\boldsymbol{a_{n-1}}=(a_{1,n-1},a_{2,n-1},a_{3,n-1},a_{4,n-1},b_{4,n-1}))\\
=&\mathbb{P}(u_{1,n}=u_{2,n-1}-a_{2,n-1}+R_{6,n},\, u_{2,n}=u_{3,n-1}-a_{3,n-1},\\
&u_{3,n}=u_{4,n-1}-a_{4,n-1}-b_{4,n-1},\, u_{4,n}=4s-(u_{1,n}+u_{2,n}+u_{3,n}))
\end{aligned}$$

$$
\begin{aligned}
&=\mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1}\, ,\, u_{2,n} = u_{3,n-1} - a_{3,n-1},\\
&\qquad u_{3,n} = u_{4,n-1} - a_{4,n-1} - b_{4,n-1}\, ,\, R_{24,n} \geq 4s - (u_{1,n} + u_{2,n} + u_{3,n}))\\
&=\mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1}\, ,\\
&\qquad\qquad R_{24,n} \geq 4s - (u_{1,n} + u_{2,n} + u_{3,n}))\\
&=\mathbb{P}(R_{6,n} = u_{1,n} - u_{2,n-1} + a_{2,n-1})\\
&\qquad \times\ \mathbb{P}(R_{24,n} \geq 4s - (u_{1,n} + u_{2,n} + u_{3,n}))\\
&=e^{-\lambda_6}\, \frac{\lambda_6^{u_{1,n}-u_{2,n-1}+a_{2,n-1}}}{(u_{1,n} - u_{2,n-1} + a_{2,n-1})!} \cdot \sum_{r_{24}=4s-(u_{1,n}+u_{2,n}+u_{3,n})}^{\infty} e^{-\lambda_{24}}\, \frac{\lambda_{24}^{r_{24}}}{r_{24}!}\\
&=e^{-\lambda_6}\, \frac{\lambda_6^{u_{1,n}-u_{2,n-1}+a_{2,n-1}}}{(u_{1,n} - u_{2,n-1} + a_{2,n-1})!} \cdot \left(1 - \sum_{r_{24}=0}^{4s-(u_{1,n}+u_{2,n}+u_{3,n})-1} e^{-\lambda_{24}}\, \frac{\lambda_{24}^{r_{24}}}{r_{24}!}\right)\\
&=Poi(u_{1,n} - u_{2,n-1} + a_{2,n-1}, \lambda_6)\, .\, F_{Erl}\left(4s - (u_{1,n} + u_{2,n} + u_{3,n}),\, \lambda_{24}\right), \qquad (4.24)\\
&u_{1,n} < s\, ;\, u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} = 4s,
\end{aligned}
$$

where $F_{Erl}(k, \lambda)$ is the cumulative distribution function of the Erlang-k distribution with scale parameter of λ.

If $u_{2,n-1} - a_{2,n-1} + R_{6,n} \geq s$, then at the end of shift n the first element of the state is $u_{1,n} = s$. Next, if $u_{1,n} + u_{2,n} + u_{3,n} + R_{24,n} < 4s$, then at the end of shift n the last element of the state is $u_{4,n} = R_{24,n}$. Hence, the state at the end of shift n is $u_{1,n} = s,\ u_{1,n} + u_{2,n} + u_{3,n} + u_{4,n} < 4s$ and the transition probability is given by:

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{u_n} \,|\, \boldsymbol{u_{n-1}}, \boldsymbol{a_{n-1}}) &=\mathbb{P}(U_{1,n} = u_{1,n},\, U_{2,n} = u_{2,n},\, U_{3,n} = u_{3,n},\, U_{4,n} = u_{4,n}|\\
&\quad U_{1,n-1} = u_{1,n-1},\, U_{2,n-1} = u_{2,n-1},\, U_{3,n-1} = u_{3,n-1},\, U_{4,n-1} = u_{4,n-1},\\
&\quad \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1}, b_{4,n-1}))\\
&=\mathbb{P}(u_{1,n} = s,\, u_{2,n} = u_{3,n-1} - a_{3,n-1},\\
&\quad u_{3,n} = u_{4,n-1} - a_{4,n-1} - b_{4,n-1},\, u_{4,n} = R_{24,n})\\
&=\mathbb{P}(s \leq u_{2,n-1} - a_{2,n-1} + R_{6,n},\, u_{2,n} = u_{3,n-1} - a_{3,n-1},\\
&\quad u_{3,n} = u_{4,n-1} - a_{4,n-1} - b_{4,n-1},\, u_{4,n} = R_{24,n})\\
&=\mathbb{P}(R_{6,n} \geq s - u_{2,n-1} + a_{2,n-1}\, ,\, R_{24,n} = u_{4,n})\\
&=\mathbb{P}(R_{6,n} \geq s - u_{2,n-1} + a_{2,n-1})\ \times\ \mathbb{P}(R_{24,n} = u_{4,n})\\
&=e^{-\lambda_{24}} \sum_{r_6=s-u_{2,n-1}+a_{2,n-1}}^{\infty} e^{-\lambda_6}\, \frac{\lambda_6^{r_6}}{r_6!} \cdot \frac{\lambda_{24}^{u_{4,n}}}{u_{4,n}!}\\
&=e^{-\lambda_{24}}\, \frac{\lambda_{24}^{u_{4,n}}}{u_{4,n}!} \left(1 - \sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}-1} e^{-\lambda_6}\, \frac{\lambda_6^{r_6}}{r_6!}\right)
\end{aligned}
$$

$$\begin{aligned}\mathbb{P}(\boldsymbol{u_n}\,|\,\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}) =& Poi(u_{4,n},\lambda_{24})\,.\,F_{Erl}\left(s-u_{2,n-1}+a_{2,n-1},\,\lambda_6\right) \\ & u_{1,n}=s\;;\;u_{1,n}+u_{2,n}+u_{3,n}+u_{4,n}<4s.\end{aligned} \tag{4.25}$$

If $u_{2,n-1}-a_{2,n-1}+R_{6,n}\geq s$, then at the end of shift n the first element of the state is $u_{1,n}=s$. Next, if $u_{1,n}+u_{2,n}+u_{3,n}+R_{24,n}\geq 4s$, then at the end of shift n the state is given by $u_{1,n}=s$, $u_{1,n}+u_{2,n}+u_{3,n}+u_{4,n}=4s$, which is equivalent to $u_{1,n}=s$, $u_{4,n}=3s-(u_{2,n}+u_{3,n})$ and the transition probability is given by:

$$\begin{aligned}\mathbb{P}(\boldsymbol{u_n}\,|\,\boldsymbol{u_{n-1}},\boldsymbol{a_{n-1}}) =&\mathbb{P}(U_{1,n}=u_{1,n},\,U_{2,n}=u_{2,n},\,U_{3,n}=u_{3,n},\,U_{4,n}=u_{4,n}| \\ & U_{1,n-1}=u_{1,n-1},\,U_{2,n-1}=u_{2,n-1},\,U_{3,n-1}=u_{3,n-1},\,U_{4,n-1}=u_{4,n-1}, \\ & \boldsymbol{a_{n-1}}=(a_{1,n-1},a_{2,n-1},a_{3,n-1},a_{4,n-1},b_{4,n-1})) \\ =&\mathbb{P}(u_{1,n}=s,\,u_{2,n}=u_{3,n-1}-a_{3,n-1}, \\ & u_{3,n}=u_{4,n-1}-a_{4,n-1}-b_{4,n-1},\,u_{4,n}=3s-(u_{2,n}+u_{3,n})) \\ =&\mathbb{P}(s\leq u_{2,n-1}-a_{2,n-1}+R_{6,n},\,u_{2,n}=u_{3,n-1}-a_{3,n-1}, \\ & u_{3,n}=u_{4,n-1}-a_{4,n-1}-b_{4,n-1},\,3s-(u_{2,n}+u_{3,n})\leq R_{24,n}) \\ =&\mathbb{P}(R_{6,n}\geq s-u_{2,n-1}+a_{2,n-1}\,,\,R_{24,n}\geq 3s-u_{2,n}-u_{3,n}) \\ =&\mathbb{P}(R_{6,n}\geq s-u_{2,n-1}+a_{2,n-1})\;\times\;\mathbb{P}(R_{24,n}\geq 3s-u_{2,n}-u_{3,n}) \\ =&e^{-\lambda_6}\,e^{-\lambda_{24}}\sum_{r_6=s-u_{2,n-1}+a_{2,n-1}}^{\infty}\frac{\lambda_6^{r_6}}{r_6!}\cdot\sum_{r_{24}=3s-u_{2,n}-u_{3,n}}^{\infty}\frac{\lambda_{24}^{r_{24}}}{r_{24}!} \\ =&\left(1-\sum_{r_6=0}^{s-u_{2,n-1}+a_{2,n-1}-1}e^{-\lambda_6}\frac{\lambda_6^{r_6}}{r_6!}\right)\cdot\left(1-\sum_{r_{24}=0}^{3s-u_{2,n}-u_{3,n}-1}e^{-\lambda_{24}}\frac{\lambda_{24}^{r_{24}}}{r_{24}!}\right) \\ =&F_{Erl}\left(s-u_{2,n-1}+a_{2,n-1},\,\lambda_6\right)\;.\,F_{Erl}\left(3s-u_{2,n}-u_{3,n},\,\lambda_{24}\right), \\ & u_{1,n}=s\;;\;u_{1,n}+u_{2,n}+u_{3,n}+u_{4,n}=4s.\end{aligned} \tag{4.26}$$

We adapt the direct cost regarding the rejected 6-hour and 24-hour patients and the unused dedicated capacity, which are shown by Equations (4.16), (4.18) and (4.19). We introduce c_b as the cost of each 24-hour patient that is deferred to another resource. Hence, for this model, the expected total costs in shift n is:

$$\mathbb{E}[c_n]=c_e\,\mathbb{E}[N_{e,n}]+\hat{c}_e\,\mathbb{E}[\hat{N}_{e,n}]+c_u\,\mathbb{E}[N_{u,n}]+c_b\,.\,b_{4,n},\;\;\forall n\in\mathbb{N}. \tag{4.27}$$

Optimality equation

We want to minimize the rejected patients, use the dedicated capacity optimally and defer the newly admitted 24-hour patients accordingly. This can be done by minimizing the expected discounted cost (using the discount factor discussed in the previous model) over an infinite horizon, as the costs are incurred from the unused dedicated capacity, rejecting patients from entering the system and deferring the newly admitted 24-hour patients to another resource. The optimality equation is then the same as discussed in the previous model (Equation 4.21). To find the optimal policy, we use the policy iteration algorithm given as Algorithm 1.

4.4 Assigning urgent patients to the appropriate shift (Model 2)

In this section we describe the MDP model to schedule urgent patients, where the action in each shift is to decide in which shifts the arriving patients are scheduled. Further elements of the model are explained in the next parts.

Decision epochs

Similar to the first model, the decision on how many shift(s) ahead the operations are scheduled, is taken at the end of each shift, which is illustrated by Figure 4.1. The notation for the decision epoch is given by:

$$\mathbb{N} = \{1, 2, \ldots\},$$
$$n \in \mathbb{N}.$$

State space

The state in the end of shift $n \in \mathbb{N}$ is the number of patients scheduled t shifts ahead along with the patient arrivals of each type. By employing Assumption 4.2.5, the definition below is given to model the state space.

Definition 4.4.1. *Let a stochastic process* $\{\mathbf{H}_n,\ n = 1, 2, 3, \ldots\}$*, where*

$$\mathbf{H}_n = \left(H_{1,n}, H_{2,n}, H_{3,n}, H_{4,n}, \bar{R}_{6,n}, \bar{R}_{24,n}\right), \forall n \in \mathbb{N}$$

and $H_{t,n}$ *records the number of urgent patients at shift* n *that are scheduled* t *shifts ahead, where* $t = 1, 2, 3, 4$, $\bar{R}_{6,n}$ *and* $\bar{R}_{24,n}$ *record the number of 6-hour and 24-hour patient admitted at shift* n*, respectively.*

Let η and $\bar{r}$ denote the realizations of the random variable H and $\bar{R}$, respectively, in Definition 4.4.1. Hence, the state at the end of shift n is denoted as:

$$\boldsymbol{\eta_n} = (\eta_{1,n}\,;\ \eta_{2,n}\,;\ \eta_{3,n}\,;\ \eta_{4,n}\,;\ \bar{r}_{6,n}\,;\ \bar{r}_{24,n}).$$

The number of patients admitted depends on the number of patient arrivals. All possible state spaces at decision epoch n build the entire state space $\mathscr{S}$, which is given by:

$$\mathscr{S} = \{S_n\}_{n\in\mathbb{N}},$$

where

$$S_n = \{\boldsymbol{\eta_n} = (\eta_{1,n}\,;\, \eta_{2,n}\,;\, \eta_{3,n}\,;\, \eta_{4,n}\,;\, \bar{r}_{6,n}\,;\, \bar{r}_{24,n})\,|\eta_{1,n}, \eta_{2,n}, \eta_{3,n}, \eta_{4,n} = 0, 1, 2, \ldots < \infty;$$
$$\bar{r}_{6,n} = \min\{R_{6,n}, s - \eta_{1,n}\}\,;\, \bar{r}_{24,n} = \min\{R_{24,n}, 4s - (\bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n})\}\}. \tag{4.28}$$

Boundary of the states

By looking at the definition of the states above and Assumption 4.2.5, the maximum number of surgeries that are scheduled up to 4 shifts ahead is $4s$, shown in the equation below:

$$\sum_{t=1}^{4} \eta_{t,n} \leq 4s,\ \forall n \in \mathbb{N}, \text{where} \tag{4.29}$$

$$\eta_{t,n} \leq s,\ t = 1, \ldots, 5,\ \forall n \in \mathbb{N}. \tag{4.30}$$

Other than that, restriction on the admitted patients is applied according to Assumption 4.3.2. The values of $\bar{r}_{6,n}$ and $\bar{r}_{24,n}$ given in Equation (4.28) depend on the dedicated capacity. Based on Assumption 4.3.2, the boundaries for the admitted patients are $0 \leq \bar{r}_{6,n} \leq s - \eta_{1,n}$ and $0 \leq \bar{r}_{24,n} \leq 4s - \left(\sum_{t=1}^{4} \eta_{t,n} + \bar{r}_{6,n}\right)$ for the admitted 6-hour and 24-hour patients, respectively. In the next part, the decision in each shift is explained.

Decisions

The decision is the number of patients scheduled in each shift up to five shifts ahead. Other than the maximum waiting time of urgent surgeries that is given by Assumption 4.2.5, the decisions should satisfy the conditions below:

1. If in shift n, the surgeries that are scheduled i shifts ahead become the surgeries that are scheduled $i-1$ shifts ahead at shift $n+1$, where $1 < i \leq 4$. For example, if in shift $n \in \mathbb{N}$ we have state $\boldsymbol{\eta_n} = (\eta_{1,n}\,;\, \eta_{2,n}\,;\, \eta_{3,n}\,;\, \eta_{4,n}\,;\, \bar{r}_{6,n}\,;\bar{r}_{24,n}) = (0, 1, 1, 1, 0, 0)$ and no arrivals at shift n, then at shift $n+1$ we have state $\boldsymbol{\eta_{n+1}} = (1, 1, 1, 0, 0, 0)$ in case of no arrivals at shift $n+1$.

2. The sum of the surgeries that are scheduled in each shift equals to the number of admitted patients in that shift.

The notation for the actions is given below.

$$A_{\boldsymbol{u},n} = \{\boldsymbol{a_n} = (a_{1,n}, a_{2,n}, a_{3,n}, a_{4,n}, a_{5,n}) \ = (\text{ schedule } a_{t,n} \text{ surgeries } t \text{ shifts ahead },$$
$$t = 1, 2, 3, 4, 5 \,|\, \sum_{t=1}^{5} a_{t,n} = \bar{r}_{6,n} + \bar{r}_{24,n}, \forall n \in \mathbb{N}\},$$
$$A = \bigcup_{\boldsymbol{u} \in \mathcal{S}, n \in \mathbb{N}} A_{\boldsymbol{u},n}.$$

Based on the states and the action in each shift, the state evolution from shift $n-1$ to shift n is given in the diagram in Figure 4.4.

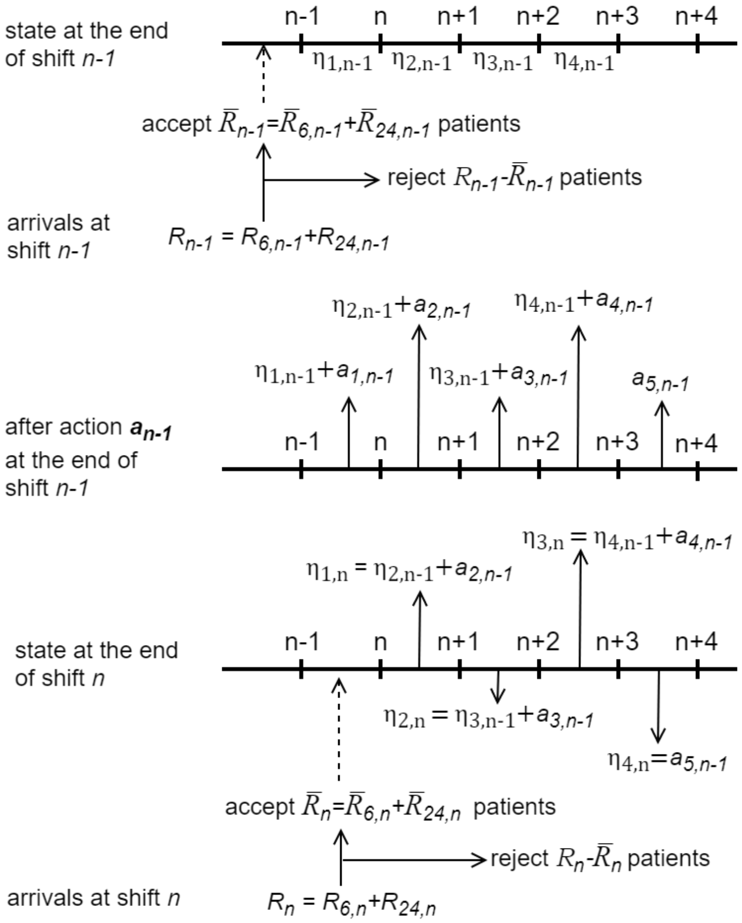


Figure 4.4: State evolution from the beginning of shift $n-1$ to the end of shift n to assign surgeries to the appropriate shifts

Transition probabilities

Considering Assumption 4.3.2 and the state evolution in Figure 4.4, we construct the transition probability. The following notation is used :

$$\begin{aligned}\mathbb{P}(\boldsymbol{\eta_n}\,|\,\boldsymbol{\eta_{n-1}},\boldsymbol{a_{n-1}}) =&\mathbb{P}(H_{1,n}=\eta_{1,n},\ H_{2,n}=\eta_{2,n},\ H_{3,n}=\eta_{3,n},\ H_{4,n}=\eta_{4,n},\ \bar{R}_{6,n}=\bar{r}_{6,n},\\ &\bar{R}_{24,n}=\bar{r}_{24,n}\,|\,H_{1,n-1}=\eta_{1,n-1},\ H_{2,n-1}=\eta_{2,n-1},\ H_{3,n-1}=\eta_{3,n-1},\\ &H_{4,n-1}=\eta_{4,n-1},\ \bar{R}_{6,n-1}=\bar{r}_{6,n-1},\ \bar{R}_{24,n-1}=\bar{r}_{24,n-1},\\ &\boldsymbol{a_{n-1}}=(a_{1,n-1},a_{2,n-1},a_{3,n-1},a_{4,n-1},a_{5,n-1}))\\ =&\mathbb{P}(\eta_{1,n}=\eta_{2,n-1}+a_{2,n-1},\ \eta_{2,n}=\eta_{3,n-1}+a_{3,n-1},\\ &\eta_{3,n}=\eta_{4,n-1}+a_{4,n-1},\eta_{4,n}=a_{5,n-1},\bar{r}_{6,n}=\min\{R_{6,n},s-\eta_{1,n}\},\\ &\bar{r}_{24,n}=\min\{R_{24,n},4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})\}).\end{aligned} \tag{4.31}$$

Based on the arrivals, if $R_{6,n} < s-\eta_{1,n}$, then at the end of shift n we have $\bar{r}_{6,n}=R_{6,n}$. After that, if $R_{24,n} < 4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})$, then $\bar{r}_{24,n}=R_{24,n}$. Hence, the state at the end of shift n is $(\eta_{1,n}\,;\,\eta_{2,n}\,;\,\eta_{3,n}\,;\,\eta_{4,n}\,;\,\bar{r}_{6,n}\,;\bar{r}_{24,n})$, where $\eta_{1,n} < s-\bar{r}_{6,n}$ and $\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+\bar{r}_{24,n} < 4s$, which transition probability is as follows:

$$\begin{aligned}\mathbb{P}(\boldsymbol{\eta_n}\,|\,\boldsymbol{\eta_{n-1}},\boldsymbol{a_{n-1}}) =&\mathbb{P}(H_{1,n}=\eta_{1,n},\ H_{2,n}=\eta_{2,n},\ H_{3,n}=\eta_{3,n},\ H_{4,n}=\eta_{4,n},\ \bar{R}_{6,n}=\bar{r}_{6,n},\\ &\bar{R}_{24,n}=\bar{r}_{24,n}\,|\,H_{1,n-1}=\eta_{1,n-1},\ H_{2,n-1}=\eta_{2,n-1},\ H_{3,n-1}=\eta_{3,n-1},\\ &H_{4,n-1}=\eta_{4,n-1},\ \bar{R}_{6,n-1}=\bar{r}_{6,n-1},\ \bar{R}_{24,n-1}=\bar{r}_{24,n-1},\\ &\boldsymbol{a_{n-1}}=(a_{1,n-1},a_{2,n-1},a_{3,n-1},a_{4,n-1},a_{5,n-1}))\\ =&\mathbb{P}(\eta_{1,n}=\eta_{2,n-1}+a_{2,n-1},\ \eta_{2,n}=\eta_{3,n-1}+a_{3,n-1},\\ &\eta_{3,n}=\eta_{4,n-1}+a_{4,n-1},\eta_{4,n}=a_{5,n-1},\bar{r}_{6,n}=R_{6,n},\bar{r}_{24,n}=R_{24,n})\\ =&\mathbb{P}(R_{6,n}=\bar{r}_{6,n}\,,\ R_{24,n}=\bar{r}_{24,n})\\ =&\mathbb{P}(R_{6,n}=\bar{r}_{6,n})\ \times\ \mathbb{P}(R_{24,n}=\bar{r}_{24,n})\\ =&e^{-\lambda_6}\,e^{-\lambda_{24}}\,\frac{\lambda_6^{\bar{r}_{6,n}}}{\bar{r}_{6,n}!}\cdot\frac{\lambda_{24}^{\bar{r}_{24,n}}}{\bar{r}_{24,n}!},\\ =&Poi(\bar{r}_{6,n},\lambda_6)\,.\,Poi(\bar{r}_{24,n},\lambda_{24})\end{aligned} \tag{4.32}$$

$$\eta_{1,n} < s-\bar{r}_{6,n}\ ;\ \bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+\bar{r}_{24,n} < 4s.$$

If $R_{6,n} < s-\eta_{1,n}$, then at the end of shift n we have $\bar{r}_{6,n}=R_{6,n}$. After that, if $4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}) \leq R_{24,n}$, then $\bar{r}_{24,n}=4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})$. Hence, the state at the end of shift n is $(\eta_{1,n}\,;\,\eta_{2,n}\,;\,\eta_{3,n}\,;\,\eta_{4,n}\,;\,\bar{r}_{6,n}\,;\bar{r}_{24,n})$, where $\eta_{1,n} < s-\bar{r}_{6,n}$ and $\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+\bar{r}_{24,n}=4s$, which transition probability is as follows:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\eta_n} \,|\, \boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}) =& \mathbb{P}(H_{1,n} = \eta_{1,n},\, H_{2,n} = \eta_{2,n},\, H_{3,n} = \eta_{3,n},\, H_{4,n} = \eta_{4,n},\, \bar{R}_{6,n} = \bar{r}_{6,n},\\
& \bar{R}_{24,n} = \bar{r}_{24,n} \,|\, H_{1,n-1} = \eta_{1,n-1},\, H_{2,n-1} = \eta_{2,n-1},\, H_{3,n-1} = \eta_{3,n-1},\\
& H_{4,n-1} = \eta_{4,n-1},\, \bar{R}_{6,n-1} = \bar{r}_{6,n-1},\, \bar{R}_{24,n-1} = \bar{r}_{24,n-1},\\
& \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1}, a_{5,n-1}))\\
=& \mathbb{P}(\eta_{1,n} = \eta_{2,n-1} + a_{2,n-1},\, \eta_{2,n} = \eta_{3,n-1} + a_{3,n-1},\\
& \eta_{3,n} = \eta_{4,n-1} + a_{4,n-1}, \eta_{4,n} = a_{5,n-1}, \bar{r}_{6,n} = R_{6,n},\\
& \bar{r}_{24,n} = 4s - (\bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n}))\\
=& \mathbb{P}(R_{6,n} = \bar{r}_{6,n}\,,\, R_{24,n} \geq 4s - (\bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n}))\\
=& \mathbb{P}(R_{6,n} = \bar{r}_{6,n})\\
& \qquad \times\, \mathbb{P}(R_{24,n} \geq 4s - (\bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n}))\\
=& e^{-\lambda_6}\, \frac{\lambda_6^{\bar{r}_{6,n}}}{\bar{r}_{6,n}!} \cdot \sum_{r_{24}=4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})}^{\infty} e^{-\lambda_{24}}\, \frac{\lambda_{24}^{r_{24}}}{r_{24}!}\\
=& e^{-\lambda_6}\, \frac{\lambda_6^{\bar{r}_{6,n}}}{\bar{r}_{6,n}!} \cdot \left(1 - \sum_{r_{24}=0}^{4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})-1} e^{-\lambda_{24}}\, \frac{\lambda_{24}^{r_{24}}}{r_{24}!}\right)\\
=& Poi(\bar{r}_{6,n}, \lambda_6)\, .\, F_{Erl}(4s - (\bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n}), \lambda_{24}) \qquad (4.33)\\
& \eta_{1,n} < s - \bar{r}_{6,n}\ ;\ \bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n} + \bar{r}_{24,n} = 4s.
\end{aligned}$$

where $F_{Erl}(k, \lambda)$ is the cumulative distribution function of the Erlang-k distribution with scale parameter of λ.

If $R_{6,n} \geq s - \eta_{1,n}$, then at the end of shift n we have $\bar{r}_{6,n} = s - \eta_{1,n}$. After that, if $R_{24,n} < 4s - (\bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n})$, then $\bar{r}_{24,n} = R_{24,n}$. Hence, the state at the end of shift n is $(\eta_{1,n}\,;\, \eta_{2,n}\,;\, \eta_{3,n}\,;\, \eta_{4,n}\,;\, \bar{r}_{6,n}\,; \bar{r}_{24,n})$, where $\eta_{1,n} = s - \bar{r}_{6,n}$ and $\bar{r}_{6,n} + \eta_{1,n} + \eta_{2,n} + \eta_{3,n} + \eta_{4,n} + \bar{r}_{24,n} < 4s$, which transition probability is as follows:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\eta_n} \,|\, \boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}) =& \mathbb{P}(H_{1,n} = \eta_{1,n},\, H_{2,n} = \eta_{2,n},\, H_{3,n} = \eta_{3,n},\, H_{4,n} = \eta_{4,n},\, \bar{R}_{6,n} = \bar{r}_{6,n},\\
& \bar{R}_{24,n} = \bar{r}_{24,n} \,|\, H_{1,n-1} = \eta_{1,n-1},\, H_{2,n-1} = \eta_{2,n-1},\, H_{3,n-1} = \eta_{3,n-1},\\
& H_{4,n-1} = \eta_{4,n-1},\, \bar{R}_{6,n-1} = \bar{r}_{6,n-1},\, \bar{R}_{24,n-1} = \bar{r}_{24,n-1},\\
& \boldsymbol{a_{n-1}} = (a_{1,n-1}, a_{2,n-1}, a_{3,n-1}, a_{4,n-1}, a_{5,n-1}))\\
=& \mathbb{P}(\eta_{1,n} = \eta_{2,n-1} + a_{2,n-1},\, \eta_{2,n} = \eta_{3,n-1} + a_{3,n-1},\\
& \eta_{3,n} = \eta_{4,n-1} + a_{4,n-1}, \eta_{4,n} = a_{5,n-1}, \bar{r}_{6,n} = s - \eta_{1,n}, \bar{r}_{24,n} = R_{24,n})\\
=& \mathbb{P}(R_{6,n} \geq s - \eta_{1,n}\,,\, R_{24,n} = \bar{r}_{24,n})
\end{aligned}$$

$$
\begin{aligned}
&=\mathbb{P}(R_{6,n} \geq s-\eta_{1,n}) \;\times\; \mathbb{P}(R_{24,n} = \bar{r}_{24,n})\\
&= \sum_{r_6=s-\eta_{1,n}}^{\infty} e^{-\lambda_6}\,\frac{\lambda_6^{r_6}}{r_6!}\,.\,e^{-\lambda_{24}}\,\frac{\lambda_{24}^{\bar{r}_{24,n}}}{\bar{r}_{24,n}!},\\
&= \left(1-\sum_{r_6=0}^{s-\eta_{1,n}-1} e^{-\lambda_6}\,\frac{\lambda_6^{r_6}}{r_6!}\right)\,.\,e^{-\lambda_{24}}\,\frac{\lambda_{24}^{\bar{r}_{24,n}}}{\bar{r}_{24,n}!},\\
&=F_{Erl}(s-\eta_{1,n},\lambda_6)\,.\,Poi(\bar{r}_{24,n},\lambda_{24}) \qquad (4.34)
\end{aligned}
$$

$$\eta_{1,n} = s-\bar{r}_{6,n}\ ;\ \bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+\bar{r}_{24,n} < 4s.$$

If $R_{6,n} \geq s-\eta_{1,n}$, then at the end of shift n we have $\bar{r}_{6,n} = s-\eta_{1,n}$. After that, if $4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}) \leq R_{24,n}$, then $\bar{r}_{24,n} = 4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})$. Hence, the state at the end of shift n is $(\eta_{1,n}\ ;\ \eta_{2,n}\ ;\ \eta_{3,n}\ ;\ \eta_{4,n}\ ;\ \bar{r}_{6,n}\ ;\bar{r}_{24,n})$, where $\eta_{1,n} = s-\bar{r}_{6,n}$ and $\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+\bar{r}_{24,n} = 4s$, which transition probability is as follows:

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{\eta_n}\,|\,\boldsymbol{\eta_{n-1}},\boldsymbol{a_{n-1}}) =&\mathbb{P}(H_{1,n}=\eta_{1,n},\ H_{2,n}=\eta_{2,n},\ H_{3,n}=\eta_{3,n},\ H_{4,n}=\eta_{4,n},\ \bar{R}_{6,n}=\bar{r}_{6,n},\\
&\bar{R}_{24,n}=\bar{r}_{24,n}\,|\,H_{1,n-1}=\eta_{1,n-1},\ H_{2,n-1}=\eta_{2,n-1},\ H_{3,n-1}=\eta_{3,n-1},\\
&H_{4,n-1}=\eta_{4,n-1},\ \bar{R}_{6,n-1}=\bar{r}_{6,n-1},\ \bar{R}_{24,n-1}=\bar{r}_{24,n-1},\\
&\boldsymbol{a_{n-1}}=(a_{1,n-1},a_{2,n-1},a_{3,n-1},a_{4,n-1},a_{5,n-1}))\\
=&\mathbb{P}(\eta_{1,n}=\eta_{2,n-1}+a_{2,n-1},\ \eta_{2,n}=\eta_{3,n-1}+a_{3,n-1},\\
&\eta_{3,n}=\eta_{4,n-1}+a_{4,n-1},\eta_{4,n}=a_{5,n-1},\bar{r}_{6,n}=s-\eta_{1,n},\\
&\bar{r}_{24,n}=4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}))\\
=&\mathbb{P}(R_{6,n}\geq s-\eta_{1,n}\ ,\ R_{24,n}\geq 4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}))\\
=&\mathbb{P}(R_{6,n}\geq s-\eta_{1,n})\\
&\qquad\times\ \mathbb{P}(R_{24,n}\geq 4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}))\\
=&\sum_{r_6=s-\eta_{1,n}}^{\infty} e^{-\lambda_6}\,\frac{\lambda_6^{r_6}}{r_6!}\,.\sum_{r_{24}=4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})}^{\infty} e^{-\lambda_{24}}\,\frac{\lambda_{24}^{r_{24}}}{r_{24}!},\\
=&\left(1-\sum_{r_6=0}^{s-\eta_{1,n}-1} e^{-\lambda_6}\,\frac{\lambda_6^{r_6}}{r_6!}\right)\\
&\qquad\left(1-\sum_{r_{24}=0}^{4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n})-1} e^{-\lambda_{24}}\,\frac{\lambda_{24}^{r_{24}}}{r_{24}!}\right)\\
=&F_{Erl}(s-\eta_{1,n},\lambda_6)\,.\,F_{Erl}(4s-(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}),\lambda_{24})\\
&\qquad (4.35)
\end{aligned}
$$

$$\eta_{1,n} = s-\bar{r}_{6,n}\ ;\ \bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+\bar{r}_{24,n} = 4s.$$

In the next part, we will address the cost corresponding to the action that we take on each shift.

Direct cost

The first cost incurred because of the rejections of the 6-hour patient arrivals from entering the system. From Assumption 4.3.2 and the boundary of the states in Equation (4.30), we know that if $R_{6,n} > s - \eta_{1,n}$, we admit $\bar{r}_{6,n} = s - \eta_{1,n}$ and reject $R_{6,n} - \bar{r}_{6,n}$ 6-hour patients. Considering $\eta_{1,n} = \eta_{2,n-1} + a_{2,n-1}$ we need to reject $R_{6,n} - \bar{r}_{6,n} = R_{6,n} - (s - \eta_{2,n-1} - a_{2,n-1})$ patients. Denoting the number of 6-hour surgeries being rejected at shift n by $N_{r,n}$, the formula to compute its expectation, $\mathbb{E}[N_{r,n}|\boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}]$, is:

$$
\begin{aligned}
\mathbb{E}[N_{r,n}|\boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}] &= \mathbb{E}[R_{6,n} - (s - \eta_{2,n-1} - a_{2,n-1})]^+ \\
&= \sum_{r_6=0}^{\infty} [(r_6 - (s - \eta_{2,n-1} - a_{2,n-1})]^+ \, \mathbb{P}(R_{6,n} = r_6) \\
&= \sum_{r_6=s-\eta_{2,n-1}-a_{2,n-1}+1}^{\infty} [r_6 - (s - \eta_{2,n-1} - a_{2,n-1})] \, \mathbb{P}(R_{6,n} = r_6) \\
&= \sum_{r_6=s-\eta_{2,n-1}-a_{2,n-1}+1}^{\infty} r_6 \, \mathbb{P}(R_{6,n} = r_6) \\
&\quad - \sum_{r_6=s-\eta_{2,n-1}-a_{2,n-1}+1}^{\infty} (s - \eta_{2,n-1} - a_{2,n-1}) \, \mathbb{P}(R_{6,n} = r_6) \\
&= \sum_{r_6=0}^{\infty} r_6 \, \mathbb{P}(R_{6,n} = r_6) - \sum_{r_6=0}^{s-\eta_{2,n-1}-a_{2,n-1}} r_6 \, \mathbb{P}(R_{6,n} = r_6) \\
&\quad - (s - \eta_{2,n-1} - a_{2,n-1}) \left(1 - \sum_{r_6=0}^{s-\eta_{2,n-1}-a_{2,n-1}} \mathbb{P}(R_{6,n} = r_6)\right) \\
&= \mathbb{E}[R_{6,n}] - \sum_{r_6=0}^{s-\eta_{2,n-1}-a_{2,n-1}} r_6 \, e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!} \\
&\quad - (s - \eta_{2,n-1} - a_{2,n-1}) \left(1 - \sum_{r_6=0}^{s-\eta_{2,n-1}-a_{2,n-1}} \mathbb{P}(R_{6,n} = r_6)\right) \\
&= \lambda_6 - \sum_{r_6=0}^{s-\eta_{2,n-1}-a_{2,n-1}} r_6 \, e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!} \\
&\quad - (s - \eta_{2,n-1} - a_{2,n-1}) \left(1 - \sum_{r_6=0}^{s-\eta_{2,n-1}-a_{2,n-1}} e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!}\right) \\
&= \lambda_6 - \lambda_6(1 - F_{Erl}(s - \eta_{2,n-1} - a_{2,n-1}, \lambda_6)) \\
&\qquad - (s - \eta_{2,n-1} - a_{2,n-1}) \left(1 - \sum_{r_6=0}^{s-\eta_{2,n-1}-a_{2,n-1}} e^{-\lambda_6} \frac{\lambda_6^{r_6}}{r_6!}\right)
\end{aligned}
$$

$$\begin{aligned}\mathbb{E}[N_{r,n}|\boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}] &= F_{Erl}(s-\eta_{2,n-1}-a_{2,n-1},\lambda_6)\\ &\qquad -(s-\eta_{2,n-1}-a_{2,n-1})\,F_{Erl}(s-\eta_{2,n-1}-a_{2,n-1},\lambda_6), \forall n\in\mathbb{N}.\end{aligned} \tag{4.36}$$

Another cost is incurred from rejecting the 24-hour surgeries because of the boundary of the state given in Equation (4.29). Denoting the number of 24-hour surgeries being rejected at shift n by $\hat{N}_{r,n}$, it is formulated as follows:

$$\hat{N}_{r,n} = (\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+R_{24,n}-4s)^{+}, \forall n\in\mathbb{N}. \tag{4.37}$$

Recall that the state at shift n depends on the state $\boldsymbol{u_{n-1}}$ and action $\boldsymbol{a_{n-1}}$. Hence, $\mathbb{E}[\hat{N}_{r,n}|\boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}]$ is formulated as follows.

$$\begin{aligned}\mathbb{E}[\hat{N}_{r,n}|\boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}] &= \mathbb{E}[(\bar{r}_{6,n}+\eta_{1,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+R_{24,n}-4s)^{+}|\boldsymbol{\eta_{n-1}}, \boldsymbol{a_{n-1}}]\\ &= \mathbb{E}[(\eta_{1,n}+R_{6,n}+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+R_{24,n}-4s)^{+}]\\ &= \sum_{r_6=0}^{\infty}\sum_{r_{24}=0}^{\infty}(\eta_{1,n}+r_6+\eta_{2,n}+\eta_{3,n}+\eta_{4,n}+r_{24}-4s)^{+}\\ &\qquad\qquad\qquad \mathbb{P}(R_{6,n}=r_6)\,\mathbb{P}(R_{24,n}=r_{24})\\ &= \sum_{r_6=0}^{s-\eta_{1,n}-1}\;\sum_{r_{24}=4s-(r_6+\sum_{t=1}^{4}\eta_{t,n})+1}^{\infty}\left(r_6+\sum_{t=1}^{4}\eta_{t,n}+r_{24}-4s\right)\\ &\qquad \mathbb{P}(R_{6,n}=r_6)\;\mathbb{P}(R_{24,n}=r_{24})\\ &\qquad +\sum_{r_6=s-\eta_{1,n}}^{\infty}\;\sum_{r_{24}=3s-\sum_{t=2}^{4}\eta_{t,n}+1}^{\infty}\left(\sum_{t=2}^{4}\eta_{t,n}+r_{24}-3s\right)\\ &\qquad \mathbb{P}(R_{6,n}=r_6)\;\mathbb{P}(R_{24,n}=r_{24})\\ &= \left(\sum_{t=1}^{4}\eta_{t,n}-4s\right)\sum_{r_6=0}^{s-\eta_{1,n}-1}\mathbb{P}(R_{6,n}=r_6)\\ &\qquad \left(1-\sum_{r_{24}=0}^{4s-(r_6+\sum_{t=1}^{4}\eta_{t,n})}\mathbb{P}(R_{24,n}=r_{24})\right)\\ &\qquad +\sum_{r_6=0}^{s-\eta_{1,n}-1}\;\sum_{r_{24}=4s-(r_6+\sum_{t=1}^{4}\eta_{t,n})+1}^{\infty}(r_6+r_{24})\\ &\qquad \mathbb{P}(R_{6,n}=r_6)\;\mathbb{P}(R_{24,n}=r_{24})\\ &\qquad +\left(\sum_{t=2}^{4}\eta_{t,n}-3s\right)\left(1-\sum_{r_6=0}^{s-\eta_{1,n}-1}\mathbb{P}(R_{6,n}=r_6)\right)\\ &\qquad \left(1-\sum_{r_{24}=0}^{3s-\sum_{t=2}^{4}\eta_{t,n}}\mathbb{P}(R_{24,n}=r_{24})\right)\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{r_6=s-\eta_{1,n}}^{\infty} \sum_{r_{24}=3s-\sum_{t=2}^{4}\eta_{t,n}+1}^{\infty} r_{24}\mathbb{P}(R_{6,n}=r_6)\,\mathbb{P}(R_{24,n}=r_{24}) \\
&= \left(\sum_{t=1}^{4}\eta_{t,n}-4s\right) \sum_{r_6=0}^{s-\eta_{1,n}-1} Poi(r_6,\lambda_6) \\
&\qquad \left(1-\sum_{r_{24}=0}^{4s-(r_6+\sum_{t=1}^{4}\eta_{t,n})} Poi(r_{24},\lambda_{24})\right) \\
&+ \sum_{r_6=0}^{s-\eta_{1,n}-1} r_6\, Poi(r_6,\lambda_6) \left(1-\sum_{r_{24}=0}^{4s-(r_6+\sum_{t=1}^{4}\eta_{t,n})} Poi(r_{24},\lambda_{24})\right) \\
&+ \sum_{r_6=0}^{s-\eta_{1,n}-1} Poi(r_6,\lambda_6) \left(\lambda_{24}-\sum_{r_{24}=0}^{4s-(r_6+\sum_{t=1}^{4}\eta_{t,n})} r_{24}Poi(r_{24},\lambda_{24})\right) \\
&\quad + \left(\sum_{t=2}^{4}\eta_{t,n}-3s\right) F_{Erl}(s-\eta_{1,n},\lambda_6)\, F_{Erl}\left(3s-\sum_{t=2}^{4}\eta_{t,n}+1,\lambda_{24}\right) \\
&\qquad\qquad + F_{Erl}(s-\eta_{1,n-1},\lambda_6)\left(\lambda_{24}-\sum_{r_{24}=0}^{3s-(\sum_{t=2}^{4}\eta_{t,n})} r_{24}Poi(r_{24},\lambda_{24})\right). \qquad (4.38)
\end{aligned}
$$

We introduce c_p as a shift-unit delay cost when a 6-hour patient is scheduled later than the deadline. This cost incurs when we admit 6-hour patients ($\bar{r}_{6,n} > 0$) and do not assign them on the next shift ($a_{1,n} < \bar{r}_{6,n}$). The delay cost is denoted by $N_{p,n}$ and formulated as follows:

$$N_{p,n} = k.\left(\bar{r}_{6,n} - a_{1,n}\right)^{+}, \qquad (4.39)$$

where k is the delay, $k = 1, 2, 3$. Also, let $\hat{c}_p$ denote the delay cost when a 24-hour patient is scheduled later than the deadline. This cost incurs when we have $\bar{r}_{24,n} > 0$ and $\sum_{t=1}^{4} a_{t,n} - \bar{r}_{6,n} < \bar{r}_{24,n}$ which is denoted by $\hat{N}_{p,n}$ and formulated as follows:

$$\hat{N}_{p,n} = \left(\bar{r}_{24,n} - \sum_{t=1}^{4} a_{t,n} + \bar{r}_{6,n}\right)^{+}. \qquad (4.40)$$

Let c_r and $\hat{c}_r$ denote the cost of each rejected 6-hour patient and each rejected 24-hour patient, respectively. Henceforth, by using Equations (4.36), (4.38), (4.39), and (4.40) as the cost components we obtain the expected total costs in shift n is:

$$\mathbb{E}[c_n] = c_r\,\mathbb{E}[N_{r,n}] + \hat{c}_r\,\mathbb{E}[\hat{N}_{r,n}] + c_p\,\mathbb{E}[N_{p,n}] + \hat{c}_p\,\mathbb{E}[\hat{N}_{p,n}], \ \ \forall n \in \mathbb{N}. \qquad (4.41)$$

Optimality equation

We want to minimize the rejected patients and schedule the urgent surgeries in time. This can be done by minimizing the expected discounted cost (using the discount factor discussed in the first model) over an infinite horizon, as the costs are incurred from rejecting patients from entering the system and delaying the urgent surgeries. The optimality equation is then the same as discussed in Model 1 (Equation 4.21). To find the optimal policy, we use the policy iteration algorithm given as Algorithm 1.

Chapter 5

Numerical experiments and results

The models shown in Chapter 4 are programmed in the 64 bits version of MATLAB on a Lenovo ThinkPad E550 with Intel(R) Core(TM) i5-5200 CPU @ 2.20GHz processor (8GB RAM) and 64 bits Windows 10. In Section 5.1, we apply Model 1 to cases that are inspired by the case of the Jeroen Bosch Ziekenhuis (JBZ), Den Bosch in 2017 and RSUP Sardjito, Indonesia in 2019.

We validate our MDP models using some methods: (1) building a Discrete Event Simulation (DES) for Model 1, (2) setting one of either 6-hour or 24-hour patient arrival rates equals 0 in the numerical experiments, (3) setting the dedicated capacity s equals 1 in the numerical experiments, (4) varying the cost combinations in the numerical experiments. From, method (2)-(4) we can check whether the optimal policy makes sense or not. Also, we can check by hand, the results of method (3), i.e., setting dedicated capacity, s, equals 1 in the numerical experiment.

5.1 Numerical experiment of Model 1

In this section, we explore the behaviour of the optimal policy as we vary the arrival rates and the maximum number of each surgery type that can be performed in the dedicated capacity. We try cases where the arrival rates are close to the ones in JBZ. The average arrival rate for the 6-hour and the 24-hour patients in JBZ are 0.54 and 1.52 patients per shift, respectively. By calculating the arrival rate per shift, we obtain that the busiest shift is between 6 AM and 12 PM with arrival rates for the 6-hour and 24-hour patients of 0.9 and 3.03 patients per shift, respectively. Next, the cost combinations used in this numerical experiment are given by the following table.

	c_e	$\hat{c}_e$	c_u
CC_1	200	100	200
CC_2	200	100	0

Table 5.1: Cost components in the numerical experiment of Model 1

The MDP provides optimal policy with the corresponding total cost. Using cost combination CC_1, the total cost and running time of each observed case are given in Table 5.2.

No	Arr. rate (patients/shift)	s	Dim	Tot.cost	Runtime (h)
1	6-hour patients : 0 24-hour patients : 0.817	1	$70 \times 70 \times 5$	$890,900$	0.02
2	6-hour patients : 0 24-hour patients : 1.82	2	$495 \times 495 \times 15$	$11,800,000$	0.62
3	6-hour patients : 0.54 24-hour patients : 1.2	2	$495 \times 495 \times 15$	$1,432,800$	0.56
4	6-hour patients : 0.54 24-hour patients : 1.52	3	$1820 \times 1820 \times 35$	$9,277,600$	6.36
5	6-hour patients : 0.9 24-hour patients : 3.03	4	$4845 \times 4845 \times 70$	$26,810,000$	40.74

*Arr. rate = arrival rate of each patient type per shift (no.of patients/shift) ; s = the maximum number of surgeries that can be performed in the dedicated capacity ; Dim = dimension of the transition probability matrix ; Tot. cost = total cost incurred ; Runtime (h) = running time of the simulation in hour.

Table 5.2: The results of the numerical experiment on Model 1

Further, in Table 5.3 part of the optimal policy for the first case is given. The rest of the optimal policy for this case is provided in Appendix B.

States				Optimal policy			
2	3	0	0	2	0	0	0
1	0	1	0	1	0	1	0

States				Optimal policy			
0	1	1	1	0	1	1	0
0	1	0	1	0	1	0	1

Table 5.3: Model1's optimal policy using CC_1 for some states where $\lambda_6 = 0.54; \lambda_{24} = 1.2$

The rest of the optimal policy for the case above is given in Appendix B. Using cost combination CC_2, we have the same optimal policy as generated using cost combination CC_1 for every case in Table 5.2. This means that when $c_e > \hat{c}_e$, the cost of the unused capacity does not affect the optimal policy, i.e., the dedicated capacity is used maximally, where the patients are treated from higher urgency levels.

We conduct a sensitivity analysis to observe the optimal policy for the cases where the arrival rates are either really close to or really far from the capacity s. Considering the length of the running time, the sensitivity analysis is performed on the case where s equals 2, using cost combination CC_1 (CC_2 result in the same optimal policy as CC_1). The cases and results are shown in the following table.

No	Arr. rate (patients/shift)	s	Dim	Tot.cost	Runtime (h)
1	6-hour patients : 0.54 24-hour patients : 1.2	2	$495 \times 495 \times 15$	$1,432,800$	0.56
4	6-hour patients : 0.54 24-hour patients : 1.45	2	$495 \times 495 \times 15$	$1,360,100$	0.52
5	6-hour patients : 1.45 24-hour patients : 0.54	2	$495 \times 495 \times 15$	$1,508,000$	0.56
6	6-hour patients : 0.05 24-hour patients : 0.2	2	$495 \times 495 \times 15$	$1,8124,00$	0.55
7	6-hour patients : 0.05 24-hour patients : 0.05	2	$495 \times 495 \times 15$	$1,820,400$	0.59

*Arr. rate = arrival rate of each patient type per shift (no.of patients/shift) ; s = the maximum number of surgeries that can be performed in the dedicated capacity ; Dim = dimension of the transition probability matrix ; Tot. cost = total cost incurred ; Runtime (h) = running time of the simulation in hour.

Table 5.4: Sensitivity analysis by varying arrival rates while $s = 2$ on Model 1

As depicted in Table 5.4, the costs get larger as the gaps between the arrival rates and the capacity s grow. This occurs due to the cost of the unused capacity because a larger gap between the arrival rate and s results in more unused capacity.

As for the optimal policies of the cases in Tables 5.2 and 5.4, they follow the same rule as the optimal policy of the first case that is given in Appendix B. The various arrival rates do not result in different optimal policies. We also conduct a numerical experiment for the arrival rates presented in Table 5.2 using cost combination CC_2 where unused capacity costs 0. The optimal policies are the same as the results from cost combination CC_1. From these results, we can conclude that the optimal policy is to treat patients in decreasing order of urgency level, until the dedicated capacity is used up or all patients are treated. First, we prove that it is optimal to treat the maximum number of the 6-hour patients in the dedicated capacity as introduced by Theorem 5.1.1.

Theorem 5.1.1. *If $c_e > \hat{c}_e$, then it is optimal to treat as many 6-hour patients as possible in the dedicated capacity in each shift $n \in \mathbb{N}$.*

Proof. Assume that at shift n we reject a 6-hour patient from using the dedicated capacity when it is not needed as the dedicated capacity is still available. This incurs a cost of c_e. Now we consider two cases.

Case 1. All 24-hour patient arrivals at shift n ($R_{24,n}$) can be treated in the rest of the dedicated capacity.
Rejecting a 6-hour patient does not make any improvements, as there are no extra 24-hour patients that can be treated in the dedicated capacity. Here, the total cost incurred is c_e.

Case 2. Some 24-hour patient arrivals at shift n ($R_{24,n}$) are rejected because of the dedicated capacity constraint.
As there are 24-hour patients that are rejected, removing a 6-hour patient from the dedicated capacity allows the system to treat an extra 24-hour patient in the dedicated capacity. This means we reject one 24-hour patient less, which saves as a cost of $\hat{c}_e$. Thus, in this case, the total cost is $c_e - \hat{c}_e$.

Whether a 6-hour patient is rejected to let a 24-hour patient being treated in the dedicated capacity or not, at the beginning of the next shift we start with the same number of surgeries in the system, formulated by $\sum_{t=2}^{4}(u_{t,n} - a_{t,n})$. Also, we incur a cost of c_e for Case 1 or $c_e - \hat{c}_e > 0$ for Case 2. By looking at the number of surgeries left in the system and the total cost incurred, we can conclude that there are no improvements made by rejecting a 6-hour patient from using the dedicated capacity. This means treating as many 6-hour patients as possible in the dedicated capacity in each shift $n \in \mathbb{N}$ is the optimal action. □

From Theorem 5.1.1, in case the 6-hour patients do not use up the dedicated capacity s ($u_{1,n} < s$), then all of the 6-hour patients are treated in the dedicated capacity ($a_{1,n} = u_{1,n} < s$). The following theorem gives further detail of the policy on performing the urgent surgeries.

Theorem 5.1.2. *If at shift n all 6-hour patients do not fully occupy the dedicated capacity s, then other surgeries are treated in the rest of the dedicated capacity, i.e., $a_{2,n} + a_{3,n} + a_{4,n} > 0$, in the decreasing order of urgency levels to use the dedicated capacity maximally.*

Proof. Assume without loss of generality, in shift n the state is $u_{t,n} > 0,\ t = 1, 2, 3, 4$. Given $u_{1,n} < s$, then based on Theorem 5.1.1 we have $a_{1,n} = u_{1,n} < s$, meaning some dedicated capacity is unused.

If $a_{2,n}+a_{3,n}+a_{4,n} = 0$, then the unused dedicated capacity is $\mathbb{E}[N_{u,n}] = s-\sum_{t=1}^{4} a_{t,n} = s - a_{1,n} > 0$. According to Equation (4.27), this causes a cost of $c_u\,\mathbb{E}[N_{u,n}] = c_u \,.\, (s - a_{1,n})$. By setting $a_{2,n} + a_{3,n} + a_{4,n} > 0$, we minimize this cost as $s - \sum_{t=1}^{4} a_{t,n} < s - a_{1,n}$

for $a_{1,n} < s$; $a_{2,n} + a_{3,n} + a_{4,n} > 0$. Other than that, not treating patients while the dedicated capacity is still available, i.e., $a_{t,n} = 0$, $t = 2, 3, 4$; causes a higher expected number of rejected 6-hour patients in the future shifts, given by the equations below.

$$\begin{aligned}\mathbb{E}[N_{e,n+1}|\boldsymbol{u_n}, \boldsymbol{a_n}] &= (u_{2,n} - s)\, F_{Erl}(s - u_{2,n} + 1, \lambda_6) + \lambda_6\, F_{Erl}(s - u_{2,n}, \lambda_6)\\ &> (u_{2,n} - a_{2,n} - s)\, F_{Erl}(s - u_{2,n} + a_{2,n} + 1, \lambda_6)\\ &\quad + \lambda_6\, F_{Erl}(s - u_{2,n} + a_{2,n}, \lambda_6), \text{where } a_{2,n} > 0. \end{aligned} \tag{5.1}$$

Next, we will prove that the patients are treated in decreasing order of urgency level.

Assume that instead of treating the patients in $u_{2,n}$, we treat those in $u_{3,n}$ or $u_{4,n}$. By looking at Equation (4.16), the expected number of the rejected 6-hour patients in shift $n + 1$ is:

$$\begin{aligned}\mathbb{E}[N_{e,n+1}|\boldsymbol{u_n}, \boldsymbol{a_n}] &= (u_{2,n} - s)\, F_{Erl}(s - u_{2,n} + 1, \lambda_6) + \lambda_6\, F_{Erl}(s - u_{2,n}, \lambda_6)\\ &> (u_{2,n} - a_{2,n} - s)\, F_{Erl}(s - u_{2,n} + a_{2,n} + 1, \lambda_6)\\ &\quad + \lambda_6\, F_{Erl}(s - u_{2,n} + a_{2,n}, \lambda_6), \text{where } a_{2,n} > 0. \end{aligned} \tag{5.2}$$

By looking at Equations (5.2), we can see that having $a_{2,n} = 0$ increases the expected number of 6-hour patient rejections in shift $n + 1$ which incurs a larger cost than having $a_{2,n} > 0$. Hence, to minimize the total cost, we should treat the maximum $a_{2,n} > 0$ out of $u_{2,n}$ patients first after treating $u_{1,n}$ patients.

Next, consider the case where we still have dedicated capacity left after treating $u_{1,n} + u_{2,n}$ surgeries, i.e., $a_{1,n} + a_{2,n} = u_{1,n} + u_{2,n} < s$. Based on Equation (4.19), to minimize $\mathbb{E}[N_{u,n}] = s - \sum_{t=1}^{4} a_{t,n}$ we set $a_{3,n} + a_{4,n} > 0$. In the next part, we prove that we treat the patients in $u_{3,n}$ first before treating those in $u_{4,n}$.

Assume that $a_{3,n} = 0$, $a_{4,n} > 0$ for every shift n. We have $u_{2,n+1} = u_{3,n}$ and $u_{3,n+1} = u_{4,n} - a_{4,n}$. Next, according to the assumption before, at shift $n + 1$ the action is $a_{3,n+1} = 0$, $a_{4,n+1} > 0$, which implies $u_{2,n+2} = u_{3,n+1}$ and $u_{3,n+2} = u_{4,n+1} - a_{4,n+1}$. The action taken at shift n and $n + 1$ affects the expected number of 6-hour patient rejections in shift $n + 2$ below:

$$\begin{aligned}\mathbb{E}[N_{e,n+2}|\boldsymbol{u_{n+1}}, \boldsymbol{a_{n+1}}] &= (u_{2,n+1} - a_{2,n+1} - s)\, F_{Erl}(s - u_{2,n+1} + a_{2,n+1} + 1, \lambda_6)\\ &\qquad + \lambda_6\, F_{Erl}(s - u_{2,n+1} + a_{2,n+1}, \lambda_6)\\ &= (u_{3,n} - a_{2,n+1} - s)\, F_{Erl}(s - u_{3,n} + a_{2,n+1} + 1, \lambda_6)\\ &\qquad + \lambda_6\, F_{Erl}(s - u_{3,n} + a_{2,n+1}, \lambda_6)\\ &> (u_{3,n} - a_{3,n} - a_{2,n+1} - s)\, F_{Erl}(s - u_{3,n} + a_{3,n} + a_{2,n+1} + 1, \lambda_6)\\ &+ \lambda_6\, F_{Erl}(s - u_{3,n} + a_{3,n} + a_{2,n+1}, \lambda_6), \text{where } a_{3,n} > 0. \end{aligned} \tag{5.3}$$

From the equation above, we can conclude that treating the patients in $u_{4,n}$ first before those in $u_{3,n}$ incurs a larger cost from rejecting the 6-hour patients from entering the system in shift $n+2$. This cost can be minimized by treating as many patients as possible out of $u_{3,n}$ patients first, because having $a_{3,n}=0$ and $a_{4,n}>0$ causes a larger cost as $c_e > \hat{c}_e$. Hence, we treat the maximum $a_{3,n}>0$ out of $u_{3,n}$ patients first after treating $u_{1,n}+u_{2,n}$ patients.

Next, consider the case where we still have dedicated capacity left after treating $u_{1,n}+u_{2,n}+u_{3,n}$ surgeries, i.e., $a_{1,n}+a_{2,n}+a_{3,n}=u_{1,n}+u_{2,n}+u_{3,n}<s$. Based on Equation (4.19), to minimize $\mathbb{E}[N_{u,n}]=s-\sum_{t=1}^{4} a_{t,n}$ we should take the action $a_{4,n}>0$. Hence, the maximum utilization of the dedicated capacity is obtained. Other than that, not treating any of the $u_{4,n}$ patients yields in the expected number of rejected 6-hour patients in shift $n+3$ as follows:

$$
\begin{aligned}
\mathbb{E}[N_{e,n+3}|\boldsymbol{u_{n+2}},\boldsymbol{a_{n+2}}] &= (u_{2,n+2}-a_{2,n+2}-s)\,F_{Erl}(s-u_{2,n+2}+a_{2,n+2}+1,\lambda_6)\\
&\qquad\qquad\qquad + \lambda_6\,F_{Erl}(s-u_{2,n+2}+a_{2,n+2},\lambda_6)\\
&= (u_{3,n+1}-a_{2,n+2}-s)\,F_{Erl}(s-u_{3,n+1}+a_{2,n+2}+1,\lambda_6)\\
&\qquad\qquad\qquad + \lambda_6\,F_{Erl}(s-u_{3,n+1}+a_{2,n+2},\lambda_6)\\
&= (u_{4,n}-a_{2,n+2}-s)\,F_{Erl}(s-u_{4,n}+a_{2,n+2}+1,\lambda_6)\\
&\qquad\qquad\qquad + \lambda_6\,F_{Erl}(s-u_{4,n}+a_{2,n+2},\lambda_6)\\
&> (u_{4,n}-a_{4,n}-a_{2,n+2}-s)\,F_{Erl}(s-u_{4,n}+a_{4,n}+a_{2,n+2}+1,\lambda_6)\\
&+ \lambda_6\,F_{Erl}(s-u_{4,n}+a_{4,n}+a_{2,n+2},\lambda_6), \text{where } a_{4,n}>0.
\end{aligned}
\tag{5.4}
$$

Hence, in every shift $n\in\mathbb{N}$ it is optimal to treat the surgeries from a higher urgency level for the purpose of using the dedicated capacity maximally. □

Remark: Even if the cost of the unused dedicated capacity is zero, Theorem 5.1.2 is still valid. This is also demonstrated by the fact that the optimal policy from CC_1 $(c_u>0)$ and CC_2 $(c_u=0)$ are the same. Further, the proof would still hold as we have the cost of rejecting 6-hour patients ($c_e>0$), regardless of the unused dedicated capacity cost. Note that neglected the cost of rejecting the 24-hour patients. Incorporating it would only strengthen the proof of the theorem.

5.2 Numerical experiment of Model 1b

The numerical experiment is conducted to observe how the optimal policy obtained with Model 1b changes for various arrival rates and dedicated capacity levels. We use the same parameters as employed in Model 1. Also, the sensitivity analysis is conducted using cost components presented in the table below.

	c_e	$\hat{c}_e$	c_u	c_b
CC_1	200	100	200	70
CC_2	200	100	200	100
CC_3	200	100	200	0
CC_4	200	100	200	200
CC_5	200	100	200	250

Table 5.5: Cost components in the numerical experiment of Model 1b

Using CC_1, the results of the numerical experiment using different arrival rates are given in Table 5.6.

No	Arr. rate (patients/shift)	s	Dim	Tot.cost	Runtime (h)
1	6-hour patients : 0 24-hour patients : 0.817	1	$70 \times 70 \times 24$	$100,790$	0.095
2	6-hour patients : 0 24-hour patients : 1.82	2	$495 \times 495 \times 129$	$1,278,500$	4.77
3	6-hour patients : 0.54 24-hour patients : 1.45	2	$495 \times 495 \times 129$	$1,438,572$	5.06
4	6-hour patients : 0.54 24-hour patients : 1.52	3	$1820 \times 1820 \times 434$	$8,490,700$	76.88

*Arr. rate = arrival rate of each patient type per shift (no.of patients/shift) ; s = the maximum number of surgeries that can be performed in the dedicated capacity ; Dim = dimension of the transition probability matrix ; Tot. cost = total cost incurred ; Runtime (h) = running time of the simulation in hour.

Table 5.6: The results of the numerical experiment on Model 1b

Using CC_1, part of the optimal policy for the third case of the table above (the total arrival rates is close to s) are given in the table below. We can see that for this case, the 24-hour patients are not always deferred even though the system is full.

States				Optimal policy				
1	1	0	1	1	1	0	0	0
0	2	0	2	0	2	0	0	0
2	4	0	2	2	0	0	0	0

States				Optimal policy				
1	1	3	3	1	1	0	0	1
2	0	1	5	2	0	0	0	1
0	0	2	6	0	0	2	0	1

Table 5.7: Model 1b's optimal policy for some states using CC_1 on Case 3

Further, employing CC_2 for the same case, some of the optimal policy are given in the following table.

States				Optimal policy				
1	1	0	1	1	1	0	0	0
0	2	0	2	0	2	0	0	0
2	4	0	2	2	0	0	0	0

States				Optimal policy				
1	1	3	3	1	1	0	0	0
2	0	1	5	2	0	0	0	0
0	0	2	6	0	0	2	0	0

Table 5.8: Model 1b's optimal policy for some states using CC_2 on Case 3

From the results above we can see that the cost combination affects the optimal policy. For example, with CC_1, in state $(0, 0, 2, 6)$ we defer one 24-hour patient, while in CC_2 we do not defer any patients. This result is intuitive as we do not defer any newly admitted 24-hour patients when the deferring cost is higher than the rejecting cost, which is the case in CC_2.

We also perform a numerical analysis using various arrival rates. Due to the running time, in the case study we use $s = 2$ with cost combination CC_1. The arrival rates, total costs and running times are shown in the following table.

No	**Arr. rate (patients/shift)**	**s**	**Dim**	**Tot.cost**	**Runtime (h)**
1	6-hour patients : 0.54 24-hour patients : 1.2	2	$495 \times 495 \times 129$	$1,531,600$	5.36
2	6-hour patients : 1.45 24-hour patients : 0.54	2	$495 \times 495 \times 129$	$2,351,300$	5.38
3	6-hour patients : 0.05 24-hour patients : 0.2	2	$495 \times 495 \times 129$	$2,203,800$	5.49
4	6-hour patients : 0.05 24-hour patients : 0.05	2	$495 \times 495 \times 129$	$2,204,300$	5.49

*Arr. rate = arrival rate of each patient type per shift (no.of patients/shift) ; s = the maximum number of surgeries that can be performed in the dedicated capacity ; Dim = dimension of the transition probability matrix ; Tot. cost = total cost incurred ; Runtime (h) = running time of the simulation in hour.

Table 5.9: Sensitivity analysis by varying arrival rates while $s = 2$ on Model 1b

Regarding the optimal policy, the first four elements of it for Model 1b are the same as the optimal policy of Model 1, given by Theorems 5.1.1 and 5.1.2, i.e., treating patients from the higher urgency levels to use the dedicated capacity maximally. Next, recall that in the previous model, having unused capacity costs 0, i.e., $c_u = 0$, results in the same optimal policy as when $c_u \neq 0$. Hence, for Model 1b, setting $c_u = 0$ in the cost combination in Table 5.5 results in the same first four elements of the optimal policy. Regarding the last element of the optimal policy, i.e., the number 24-hour patients being deferred, based on the numerical experiment results we propose the following theorem.

Theorem 5.2.1. *If $\hat{c}_e < c_b$, then it is optimal not to defer any 24-hour patients in each shift $n \in \mathbb{N}$.*

Proof. Assume that at shift n a 24-hour patient is deferred when the dedicated capacity is still available. This causes a cost of c_b. Note that deferring a 24-hour patient at shift n affects the number of 24-hour patients admitted at shift $n+1$. Now we consider two possible events at shift $n+1$.

Case 1. All 24-hour arrivals can be admitted to the system.
In this case, deferring a 24-hour patient at shift n does not make any improvement as no extra patients can be admitted. Hence, the total cost is c_b.

Case 2. Some 24-hour arrivals are rejected.
As there are rejected 24-hour arrivals at shift $n+1$, deferring a 24-hour patient at shift n allows one extra 24-hour patient to be admitted at shift $n+1$. This saves us a cost of $\hat{c}_e$. Hence, the total cost is $c_b - \hat{c}_e > 0$.

From the cases above, we can conclude that when $\hat{c}_e < c_b$, it is optimal not to defer 24-hour patients. □

We present the optimal policy for the case where $\lambda_6 = 0, \lambda_{24} = 0.817, s = 1$ using CC_1 in Appendix C. We also give the limiting distribution of the system that results from the optimal policy for some of the studied cases above in Appendix E.2.

5.3 Numerical experiment of Model 2

In this section, we present the numerical experiment of Model 2 using a cost combination where $c_e = 200\,;\ \hat{c}_e = 100\,;\ c_p = 200\,;\ \hat{c}_p = 200$. The arrival rates and dedicated capacity along with the results are shown in the table below.

No	Arr. rate (patients/shift)	s	Dim	Tot.cost	Runtime (h)
1	6-hour patients : 0.45 24-hour patients : 0.5	1	$80 \times 80 \times 31$	$65,231$	0.11
2	6-hour patients : 0 24-hour patients : 0.817	1	$80 \times 80 \times 31$	$25,184$	0.12
3	6-hour patients : 0 24-hour patients : 1.82	2	$973 \times 973 \times 237$	$798,010$	15.55
4	6-hour patients : 0.54 24-hour patients : 1.2	2	$973 \times 973 \times 237$	$851,590$	15.9

5	6-hour patients : 0.54 24-hour patients : 1.45	2	$973 \times 973 \times 237$	$1,007,200$	15.99
6	6-hour patients : 1.45 24-hour patients : 0.54	2	$973 \times 973 \times 237$	$1,017,300$	15.6

*Arr. rate = arrival rate of each patient type per shift (no.of patients/shift) ; s = the maximum number of surgeries that can be performed in the dedicated capacity ; Dim = dimension of the transition probability matrix ; Tot. cost = total cost incurred ; Runtime (h) = running time of the simulation in hour.

Table 5.10: The results of the numerical experiment on Model 2

Part of the optimal policy when $\lambda_6 = 0.54; \lambda_{24} = 1.45; s = 2$ is presented in the following table.

States						Optimal policy					States						Optimal policy				
1	1	0	0	1	0	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0
0	1	0	2	1	0	1	0	0	0	0	0	1	2	1	0	1	1	0	0	0	0
0	1	2	1	2	0	2	0	0	0	0	0	1	2	2	2	1	2	1	0	0	0

Table 5.11: Model 2's optimal policy for some states when $\lambda_6 = 0.54; \lambda_{24} = 1.45; s = 2$

For this model, when there is 6-hour patients waiting, the optimal policy is to schedule them in the earliest available capacity. Meanwhile, there is no strict rule to schedule the 24-hour patients as long as they are assigned within the deadline.

In Model 2, the numerical experiments take longer (in running time) than those in the first two models, due to the larger state space. For this reason, we do not include the case where $\lambda_6 = 0.54; \lambda_{24} = 1.52; s = 3$ that have been observed using the two previous models.

Next, a sensitivity analysis is conducted by varying the arrival rates. Due to the running time, the dedicated capacity s of 1 is employed in this sensitivity analysis. Using the same cost combination as given in the beginning of this section, the following table shows the results of the sensitivity analysis.

No	Arr. rate (patients/shift)	s	Dim	Tot.cost	Runtime (h)
1	6-hour patients : 0.45 24-hour patients : 0.5	1	$80 \times 80 \times 31$	$65,231$	0.11
2	6-hour patients : 0.05 24-hour patients : 0.05	1	$80 \times 80 \times 31$	$17,470$	0.15

3	6-hour patients : 0.05 24-hour patients : 0	1	$80 \times 80 \times 31$	$17,243$	0.13
4	6-hour patients : 0.05 24-hour patients : 0.94	1	$80 \times 80 \times 31$	$40,195$	0.13

*Arr. rate = arrival rate of each patient type per shift (no.of patients/shift) ; s = the maximum number of surgeries that can be performed in the dedicated capacity ; Dim = dimension of the transition probability matrix ; Tot. cost = total cost incurred ; Runtime (h) = running time of the simulation in hour.

Table 5.12: Sensitivity analysis by varying arrival rates while $s = 1$ in Model 2

We also perform a case study where $c_p > \hat{c}_p$, which results in the same optimal policy compared to when $c_p = \hat{c}_p$. Next, in scheduling the 6-hour patients the next two theorems describe the optimal policy.

Theorem 5.3.1. *If $c_e > \hat{c}_e$, then it is optimal to treat as many 6-hour patients as possible in the dedicated capacity at each shift $n \in \mathbb{N}$.*

Proof. Assume that at shift n a 6-hour patient is rejected when it is not needed as the dedicated capacity is still available. This decision incurs a cost of c_e. Next, there are two possible events for the 24-hour arrivals at shift n:

Case 1. All 24-hour patient arrivals can be treated in the dedicated capacity.
In this case, rejecting 6-hour patients makes no improvement as no extra 24-hour patient can be admitted to the dedicated capacity system. Hence, in Case 1, the total cost is c_e.

Case 2. Some 24-hour arrivals are rejected because of capacity constraint.
In this case, rejecting a 6-hour patient allows an extra 24-hour patient to be admitted in the dedicated capacity. This saves us a cost of $\hat{c}_e$. Hence, for this case, the total cost incurred is $c_e - \hat{c}_e > 0$.

Further, whether a 6-hour patient is rejected or not, in the next shift all admitted patients are scheduled and we end up with the same number of patients waiting. Thus, no improvements are made by rejecting a 6-hour patient when it is not needed. □

Theorem 5.3.2. *If $c_p > 0$, then it is optimal to schedule the 6-hour patients as soon as possible.*

Proof. In this model, when there are 6-hour patients waiting, the number of patients that are scheduled on the next shift $(\eta_{1,n})$ is less than the capacity s (Equation (4.28)). This also means that some dedicated capacity is available on the next shift. Delaying a 6-hour patient at shift n by k shifts costs $k \cdot c_p, \;\; k = 1, 2, 3.$ In case a 24-hour patient is waiting, this allows us to treat him in the dedicated capacity that was initially assigned to the 6-hour patient. However, this action does not save us any costs.

Take $k = 1.$ In shift $n + 1,$ if all arrivals can be admitted, then no improvements are made. However, if in shift $n+1$ an arrival is rejected because of the delay, then we incur a cost of c_e (for a rejected 6-hour patient) or $\hat{c}_e$ (for a rejected 24-hour patient).

Similar with the previous case, for $2 \leq k \leq 3$, if all arrivals can be admitted in shift $n+1$ up to $n+k$, then there is no improvement. However, if in any shift within interval of $n+1$ up to $n+k$ we reject an arrival, then it costs us c_e (for a rejected 6-hour patient) or $\hat{c}_e$ (for a rejected 24-hour patient) on the shift where the rejection take place.

Hence, delaying a 6-hour patient by k shifts costs us $k \cdot c_p,\ \ c_p \neq 0,\ \ k = 1, 2, 3,$ while in the next k shifts no extra patients can be admitted and even a cost of at least c_e or $\hat{c}_e$ may be incurred. □

Further, by looking at the delay cost for the 24-hour patient, the following theorem gives the optimal policy in scheduling the 24-hour patients.

Theorem 5.3.3. *if $\hat{c}_p > 0$, then it is optimal to have none of the 24-hour patients delayed.*

Proof. In this model, the total number of patients in the system (the scheduled and newly admitted ones) up to 4 shifts ahead is $4s$ (Equation (4.30)). Hence, all admitted patients can be treated in the dedicated capacity within 4 shifts.

Assume that a 24-hour patient is delayed, i.e., treated 5 shifts ahead upon its arrival. It incurs a cost of $\hat{c}_p$. In case in the next shift, up to 4 shift ahead no 24-hour arrivals are rejected, then the total cost is $\hat{c}_p$. However, if in any of the shifts from the next shift up to 4 shifts ahead an arrival is rejected, then in the corresponding shift a cost of c_e (for rejecting a 6-hour patient) or $\hat{c}_e$ (for rejecting a 24-hour patient) is incurred. □

We also analyze a case study where $c_e = \hat{c}_e = 200$, while other cost components stay the same. It is interesting to see that the 24-hour patients are scheduled in the fastest available capacity after assigning the 6-hour patients. This matches with the intuition as in this case there is no preference in rejection (initially is given by the different rejection costs). Hence, the model does not try to prevent the 6-hour rejections by scheduling the 24-hour patient to the later shifts (within the target time) anymore.

In the next part, we present the fraction of time where the patients are rejected in each MDP model.

5.4 Fraction of time patients are rejected

In this section we present the fraction of time where patients are rejected in each MDP model. Recall that in Model 1 and 1b, the maximum number of patients that can wait up to the next shift is s and the maximum number of patients in the system is $4s$. Based on this, if the first element of the state is at least s, then the 6-hour patients are rejected. Also, if the total number of patients in the system is $4s,$ the 24-hour patients are rejected. Meanwhile, in Model 2, the newly arriving 6-hour patients plus the patients scheduled in the next shift should not exceed $s.$ Also, the 24-hour patients are admitted such that the total number of patients (the scheduled and newly admitted) in the system is a maximum of $4s.$

First, we look at the case where the total arrival rate is much less from the dedicated capacity: $\lambda_6 = 0.05; \lambda_{24} = 0.05; s = 1.$ Using Model 1 and 1b and CC_1 of each model, the fraction of time where we reject 6-hour and 24-hour patients are 0.049 and 2.055×10^{-6}, respectively, for both models. The results from Model 1 and 1b are the same as the first four elements of the optimal policy provided by Model 1b are the same as those from Model 1 and in Model 1b we do not defer any 24-hour patients as the dedicated capacity can handle arrivals without deferring any 24-hour patients, also as $\hat{c}_e > c_b$ in CC_1 (Theorem 5.2.1). Also, intuitively, as the total arrival rate is much less than s, we can handle all arrivals in the dedicated capacity which results in no deferred 24-hour patients. For the same total arrival rates, using Model 2 with $c_e > \hat{c}_e$ (same as in CC_1 of the previous models), the fraction of time where we reject 6-hour and 24-hour patients are 0.052 and 2.421×10^{-6}, respectively.

Next, we observe the case where the total arrival rate is close to the dedicated capacity: $\lambda_6 = 0.54; \lambda_{24} = 1.45; s = 2.$ Using CC_1 in Model 1, the 6-hour and 24-hour patients are rejected 22.8% and 9% of the time, respectively. Next, using CC_1 in Model 1b result in rejecting 6-hour and 24-hour patients 14% and 1.6% of the time, respectively. For the same amount of total arrival rate, using Model 2 with $c_e > \hat{c}_e$ (same as in CC_1 of the previous models), we reject 6-hour and 24-hour patients 23.3% and 6.4% of the time, respectively.

When the total arrival rate is close to the dedicated capacity level (s), employing

Model 1b results in less fraction of time rejecting patients than when we use Model 1. This is because in Model 1b we defer 24-hour patients to another resource when the system is full (considering in CC_1 the cost of rejecting 24-hour patients is higher than the cost of deferring newly admitted 24-hour patients). Recall that the goal of Model 1b is to reject less 6-hour patients by deferring 24-hour patients beforehand. By looking at the fraction of time where the 6-hour patients are rejected, this goal is achieved as in Model 1b the 6-hour patient rejection is decreased by 9% compared to when we use Model 1.

In Model 2, we inform urgent patients immediately when they are scheduled. However, compared to Model 1, employing Model 2 has around 1% higher fraction of the time where the 6-hour patients are rejected (same results for both cases where the total arrival rate is much less than and close to the dedicated capacity level, s).

In the Appendix D the limiting probability of for some other cases of each model is presented.

Chapter 6

Conclusions and recommendations

In this chapter, we present the conclusions and future research on the topic we discussed in the previous chapters. First, the conclusions are presented in Section 6.1. Then Section 6.2 consists of our recommendations.

6.1 Conclusions

In determining the dedicated capacity, when we have arrivals of both 6-hour and 24-hour patients (as in the JBZ case), we can compare the results from $M/M/1/4s$ and $M/M/1$ with priority queueing models. In the $M/M/1/4s$ queueing model, we observe that the 6-hour patients do not have priority over the 24-hour patients as all patients are treated according to the First-Come First-Served (FCFS) rule. In the $M/M/1$ with priority queueing model, the 6-hour patients have priority over the 24-hour patients. The models provide some variables to be considered when hospitals decide the amount of dedicated capacity. In general, higher operating room (OR) utilization results in longer waiting times. Using the results of the queueing models, OR managers can balance the OR utilization and patients' waiting time according to their preferences or the hospital guidelines. Besides, when the dedicated capacity level has the closest value to the total arrival rate, on average the 6-hour patients wait longer in $M/M/1/4s$ model than in $M/M/1$ with priority queueing model because of the priority rule. For the same case, due to the priority rule, the 24-hour patients wait longer in $M/M/1$ with priority queueing model than in $M/M/1/4s$ queueing model. Hence, in this case, the OR manager can decide upon whether to give the 6-hour patients priority (using $M/M/1$ with priority queueing model) or not (using $M/M/1/4s$ queueing model). For the case where s is much larger than the total arrival rate, giving the 6-hour patients priority over the 24-hour patients does not matter. This is because in this case, using any of the two queueing models results in the same patients' waiting times. However, maintaining such a large dedicated capacity results in a low OR utilization (high idle time).

After determining the dedicated capacity level, we proposed three Markov decision processes (MDP) based models to schedule the urgent patients. In these models, the 6-hour patients can afford to wait up to 1 shift ahead, while the 24-hour can wait up to 4 shifts ahead. In the first MDP model, in each shift we determine which patients waiting will get their surgery in the next shift. This way, the model provides a high flexibility for the hospital to schedule patients. This is because we keep track of the deadline of each patient and the MDP model results in optimal policy to schedule the urgent patients from the earliest deadlines to the later ones. The 6-hour patients that arrive later than the waiting 24-hour patients can be treated earlier by considering their deadlines. This implies that sometimes the 24-hour patients have to stay sober for a long time, because they are treated when the dedicated capacity is still available after treating the 6-hour patients, which depends on the arriving 6-hour patients.

In the MDP Model 1b, we modify the first model such that the 24-hour patients can be deferred to another resource. The 24-hour patients may be deferred from using the dedicated capacity to let extra patients enter the system. This decision depends on the costs. Similar to the first model, patients are treated from the earlier deadlines to the later ones. Thus, this model provides a higher flexibility for the hospital to schedule the 6-hour patients. In practice, deferring the 24-hour patients may cause elective cancellations, which is represented by the deferring cost in our model.

Note that the 6-hour patients are more important than the 24-hour patients as finding alternative capacity is more difficult and diverting them to another resource results in a higher medical risk. In the last MDP model (Model 2), we keep track of the number of urgent patients that are scheduled as well as the number of new urgent patient admissions. We decide the schedule of the new admissions as soon as they arrive based on their deadlines. This way the hospital does not keep the patients sober for an unnecessary period of time. However, the hospital looses its flexibility in scheduling urgent patients because in this model we reject 6-hour patients (and may still be able to admit 24-hour patients) when the schedule on the next shift is full, despite of the fact that the 6-hour patients are more important than the 24-hour patients.

In handling the total arrival rate that is much less than the dedicated capacity level, our numerical experiments show that Model 1 and 1b result in the same fraction of time where urgent patients are rejected. Meanwhile, for the case where the total arrival rate is close to the dedicated capacity level, the urgent patients are less often rejected in Model 1b than in Model 1. For Model 2, the fraction of time where the 6-hour patients are rejected is around 1% higher than the result from Model 1.

6.2 Recommendations

We recommend hospitals to use the queueing models to determine the optimal dedicated capacity levels. The hospitals can observe the system utilization, mean waiting times and queue length of the urgent patients for various dedicated capacity level using an $M/M/1/4s$ queueing model (if urgent patients are treated according to FCFS discipline) and an $M/M/1$ with priority queueing model (if urgent patients are treated using the priority rule). Next, assuming that the 30-minute patients are always scheduled in the shift where they arrive, it is recommended to schedule patients in the order of decreasing urgency levels (from Model 1) and defer newly admitted 24-hour patients when the system is full (depends on the costs; from Model 1b). Another recommendation is to schedule the 6-hour patients as soon as possible and schedule the 24-hour patients in time (from Model 2).

In future research, a Discrete Event Simulation (DES) of the models can be developed to observe the behaviour of the system when the optimal policy from the MDP model is applied. Further for practical use, some heuristics can be developed based on our MDP models and the DES. Other than that, Model 1b can be modified by not only deferring the newly admitted 24-hour patients, but also the patients who can afford to wait up to t shifts ahead, $t = 1, 2, 3$. As this modification may result in a high-dimensional action, the appropriate algorithm to solve the MDP should be studied further. Hyeong Soo et al. [22] provide some alternatives algorithms to solve this kind of MDP problem. Also, using an Approximate Dynamic Programming (ADP) approach, cancelling elective patients to schedule the urgent patients can be incorporated. ADP is used to tackle the high-dimensional state that is required to model the dynamic capacity that results from cancelling the elective patients (which is too complicated for the MDP model).

Bibliography

[1] C. Van Riet and E. Demeulemeester, "Trade-offs in operating room planning for electives and emergencies: a review," *Operations Research for Health Care*, vol. 7, pp. 52–69, 2015.

[2] E. R. Tsai, "Optimal time allocation of an orthopedic surgeon," Master's thesis, University of Twente, August 2017.

[3] D. I. McIsaac, K. Abdulla, H. Yang, S. Sundaresan, P. Doering, S. G. Vaswani, K. Thavorn, and A. J. Forster, "Association of delay of urgent or emergency surgery with mortality and use of health care resources: a propensity score–matched observational cohort study," *Canadian Medical Association Journal*, vol. 189, no. 27, pp. E905–E912, 2017.

[4] A. Macario, T. Vitez, B. Dunn, and T. McDonald, "Where are the costs in perioperative care?: Analysis of hospital costs and charges for inpatient surgical care," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 83, no. 6, pp. 1138–1144, 1995.

[5] B. Cardoen, E. Demeulemeester, and J. Beliën, "Operating room planning and scheduling: A literature review," *European journal of operational research*, vol. 201, no. 3, pp. 921–932, 2010.

[6] F. Guerriero and R. Guido, "Operational research in the management of the operating theatre: a survey," *Health care management science*, vol. 14, no. 1, pp. 89–114, 2011.

[7] M. E. Zonderland, R. J. Boucherie, N. Litvak, and C. L. Vleggeert-Lankamp, "Planning and scheduling of semi-urgent surgeries," *Health Care Management Science*, vol. 13, no. 3, pp. 256–267, 2010.

[8] P. J. Hulshof, N. Kortbeek, R. J. Boucherie, E. W. Hans, and P. J. Bakker, "Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS," *Health systems*, vol. 1, no. 2, pp. 129–175, 2012.

[9] D. Gupta, “Surgical suites’ operations management,” *Production and Operations Management*, vol. 16, no. 6, pp. 689–700, 2007.

[10] E. W. Hans, M. Van Houdenhoven, and P. J. Hulshof, “A framework for healthcare planning and control,” in *Handbook of healthcare system scheduling*. Springer, 2012, pp. 303–320.

[11] Y. Ferrand, M. Magazine, and U. Rao, “Comparing two operating-room-allocation policies for elective and emergency surgeries,” in *Proceedings of the Winter Simulation Conference*. Winter Simulation Conference, 2010, pp. 2364–2374.

[12] Y. Gerchak, D. Gupta, and M. Henig, “Reservation planning for elective surgery under uncertain demand for emergency surgery,” *Management Science*, vol. 42, no. 3, pp. 321–334, 1996.

[13] J. T. van Essen, E. W. Hans, J. L. Hurink, and A. Oversberg, “Minimizing the waiting time for emergency surgery,” *Operations Research for Health Care*, vol. 1, no. 2-3, pp. 34–44, 2012.

[14] J. Zhou and F. Dexter, “Method to assist in the scheduling of add-on surgical cases-upper prediction bounds for surgical case durations based on the log-normal distribution,” *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 89, no. 5, pp. 1228–1232, 1998.

[15] J. Beliën, E. Demeulemeester, and B. Cardoen, “A decision support system for cyclic master surgery scheduling with multiple objectives,” *Journal of scheduling*, vol. 12, no. 2, p. 147, 2009.

[16] F. Dexter, A. Macario, and R. D. Traub, “Optimal sequencing of urgent surgical cases,” *Journal of Clinical Monitoring and Computing*, vol. 15, no. 3/4, pp. 153–162, 1999.

[17] E. Erdem, X. Qu, and J. Shi, “Rescheduling of elective patients upon the arrival of emergency patients,” *Decision Support Systems*, vol. 54, no. 1, pp. 551–563, 2012.

[18] J. F. Shortle, J. M. Thompson, D. Gross, and C. M. Harris, *Fundamentals of queueing theory*. John Wiley & Sons, 2018, vol. 399.

[19] I. Adan and J. Resing, “Queueing theory,” 2002.

[20] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2005.

[21] S. M. Ross, *Introduction to probability models*. Academic press, 2014.

[22] H. S. Chang, M. C. Fu, J. Hu, S. I. Marcus *et al.*, “A survey of some simulation-based algorithms for markov decision processes,” *Communications in Information & Systems*, vol. 7, no. 1, pp. 59–92, 2007.

Appendix A

Sensitivity analysis on the $M/M/1$ with priority queueing model

In this appendix, we present the sensitivity analysis results for Case 2 and 3 of Table 3.2. In Figure A.1, the server utilization and expected waiting time for $\lambda_6 = 0.54$; $\lambda_{24} = 1.52$ on $M/M/1$ with priority rule queueing model is shown.

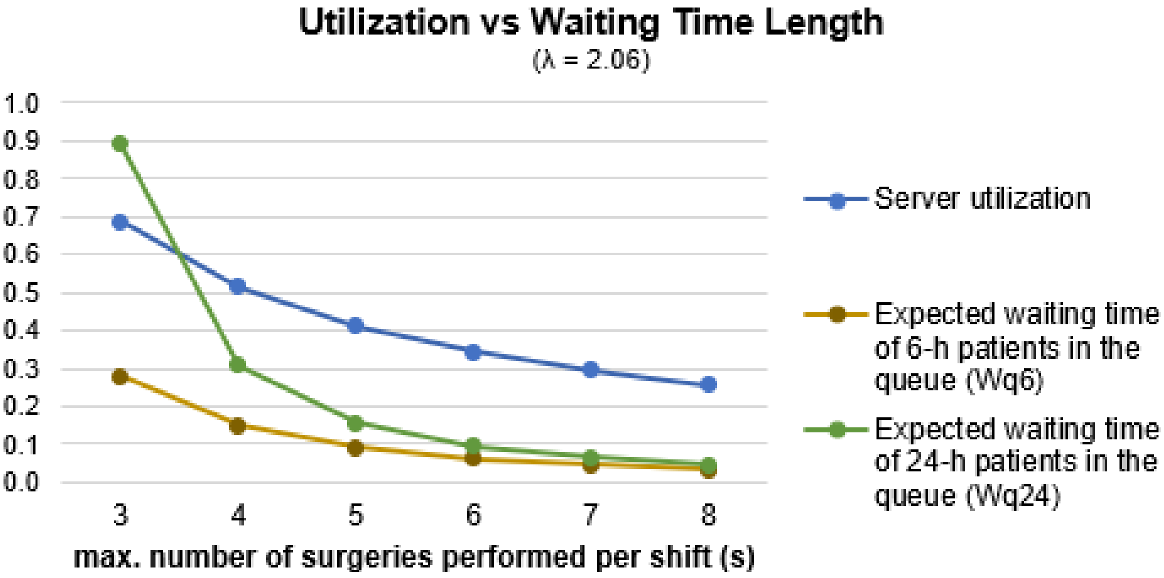


Figure A.1: Sensitivity analysis on the mean waiting time $\lambda_6 = 0.54$; $\lambda_{24} = 1.52$ using priority rule

Next, Figure A.2 represents the server utilization and expected queue length for this case.

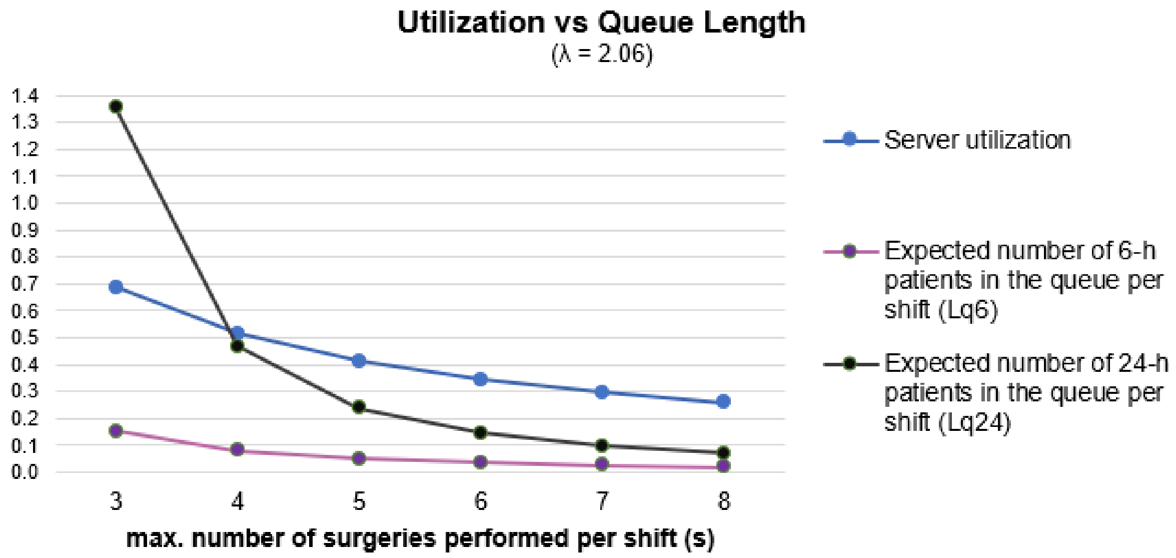


Figure A.2: Sensitivity analysis on the mean queue length $\lambda_6 = 0.54$; $\lambda_{24} = 1.52$ using priority rule

Next, we present the sensitivity analysis results for Case 3 $\lambda_6 = 0.9$; $\lambda_{24} = 3.03$ in the following figures.

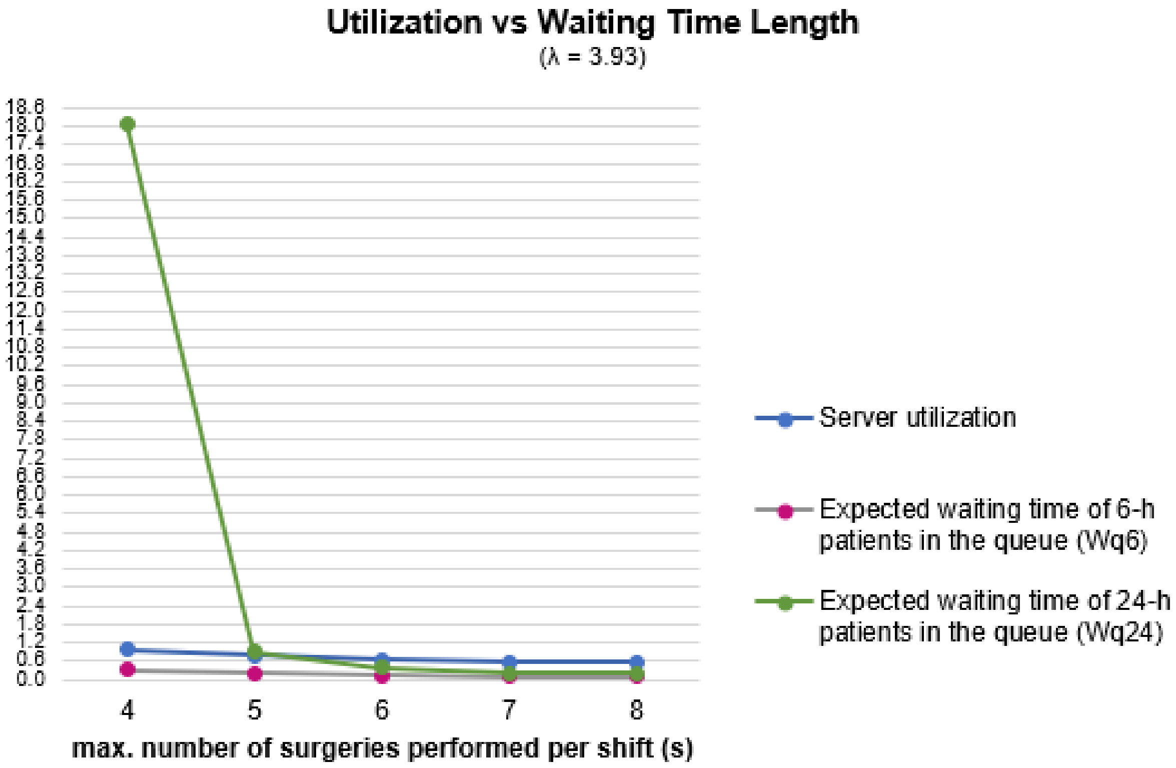


Figure A.3: Sensitivity analysis on the mean waiting time $\lambda_6 = 0.9$; $\lambda_{24} = 3.03$ using priority rule

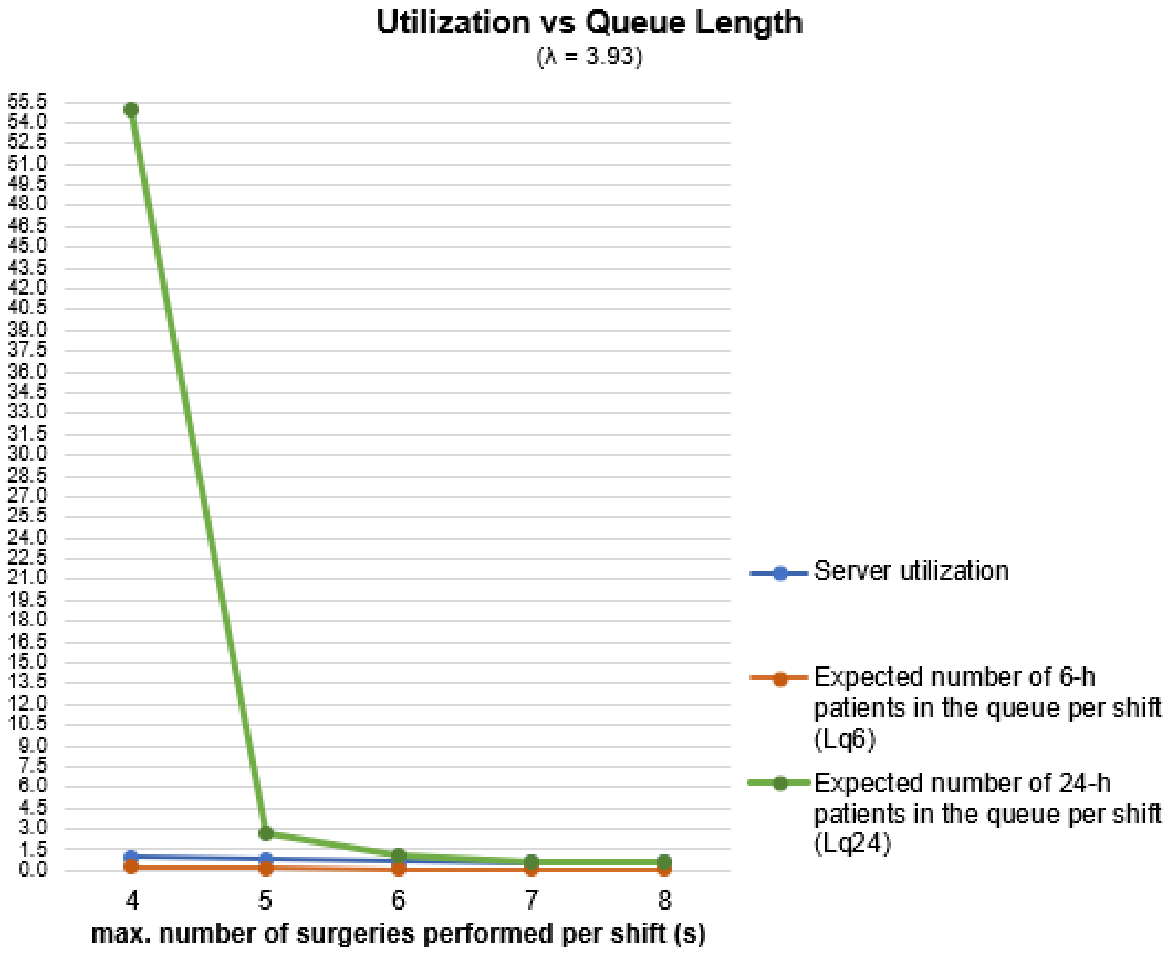


Figure A.4: Sensitivity analysis on the mean queue length $\lambda_6 = 0.9$; $\lambda_{24} = 3.03$ using priority rule

We can see that higher utilization results in longer waiting times and queues. The 24-hour patients wait longer than the 6-hour patients due to the priority rule. Also, because of the priority rule, there are more 24-hour patients than the 6-hour patients in the queue.

Appendix B

Optimal policy of Model 1

In this appendix, the optimal policy for Model 1 when $\lambda_6 = 0.54, \lambda_{24} = 1.2,\ s = 2$ using CC_1 (of Model 1) is shown.

Table B.1: Optimal policy Model 1 when $\lambda_6 = 0.54, \lambda_{24} = 1.2,\ s = 2$ using CC_1

STATES				OPTIMAL POLICY			
0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0
2	0	0	0	2	0	0	0
3	0	0	0	2	0	0	0
4	0	0	0	2	0	0	0
5	0	0	0	2	0	0	0
6	0	0	0	2	0	0	0
7	0	0	0	2	0	0	0
8	0	0	0	2	0	0	0
0	1	0	0	0	1	0	0
1	1	0	0	1	1	0	0
2	1	0	0	2	0	0	0
3	1	0	0	2	0	0	0
4	1	0	0	2	0	0	0
5	1	0	0	2	0	0	0
6	1	0	0	2	0	0	0
7	1	0	0	2	0	0	0
0	2	0	0	0	2	0	0
1	2	0	0	1	1	0	0
2	2	0	0	2	0	0	0
3	2	0	0	2	0	0	0
4	2	0	0	2	0	0	0
5	2	0	0	2	0	0	0
6	2	0	0	2	0	0	0
0	3	0	0	0	2	0	0
1	3	0	0	1	1	0	0
2	3	0	0	2	0	0	0
3	3	0	0	2	0	0	0
4	3	0	0	2	0	0	0
5	3	0	0	2	0	0	0
0	4	0	0	0	2	0	0
1	4	0	0	1	1	0	0
2	4	0	0	2	0	0	0
3	4	0	0	2	0	0	0
4	4	0	0	2	0	0	0
0	5	0	0	0	2	0	0
1	5	0	0	1	1	0	0
2	5	0	0	2	0	0	0
3	5	0	0	2	0	0	0
0	6	0	0	0	2	0	0
1	6	0	0	1	1	0	0
2	6	0	0	2	0	0	0
0	7	0	0	0	2	0	0
1	7	0	0	1	1	0	0
0	8	0	0	0	2	0	0
0	0	1	0	0	0	1	0
1	0	1	0	1	0	1	0
2	0	1	0	2	0	0	0
3	0	1	0	2	0	0	0
4	0	1	0	2	0	0	0
5	0	1	0	2	0	0	0
6	0	1	0	2	0	0	0
7	0	1	0	2	0	0	0
0	1	1	0	0	1	1	0
1	1	1	0	1	1	0	0
2	1	1	0	2	0	0	0
3	1	1	0	2	0	0	0
4	1	1	0	2	0	0	0

STATES				OPTIMAL POLICY			
5	1	1	0	2	0	0	0
6	1	1	0	2	0	0	0
0	2	1	0	0	2	0	0
1	2	1	0	1	1	0	0
2	2	1	0	2	0	0	0
3	2	1	0	2	0	0	0
4	2	1	0	2	0	0	0
5	2	1	0	2	0	0	0
0	3	1	0	0	2	0	0
1	3	1	0	1	1	0	0
2	3	1	0	2	0	0	0
3	3	1	0	2	0	0	0
4	3	1	0	2	0	0	0
0	4	1	0	0	2	0	0
1	4	1	0	1	1	0	0
2	4	1	0	2	0	0	0
3	4	1	0	2	0	0	0
0	5	1	0	0	2	0	0
1	5	1	0	1	1	0	0
2	5	1	0	2	0	0	0
0	6	1	0	0	2	0	0
1	6	1	0	1	1	0	0
0	7	1	0	0	2	0	0
0	0	2	0	0	0	2	0
1	0	2	0	1	0	1	0
2	0	2	0	2	0	0	0
3	0	2	0	2	0	0	0
4	0	2	0	2	0	0	0
5	0	2	0	2	0	0	0
6	0	2	0	2	0	0	0
0	1	2	0	0	1	1	0
1	1	2	0	1	1	0	0
2	1	2	0	2	0	0	0
3	1	2	0	2	0	0	0
4	1	2	0	2	0	0	0
5	1	2	0	2	0	0	0
0	2	2	0	0	2	0	0
1	2	2	0	1	1	0	0
2	2	2	0	2	0	0	0
3	2	2	0	2	0	0	0
4	2	2	0	2	0	0	0
0	3	2	0	0	2	0	0
1	3	2	0	1	1	0	0
2	3	2	0	2	0	0	0
3	3	2	0	2	0	0	0
0	4	2	0	0	2	0	0
1	4	2	0	1	1	0	0
2	4	2	0	2	0	0	0
0	5	2	0	0	2	0	0
1	5	2	0	1	1	0	0
0	6	2	0	0	2	0	0
0	0	3	0	0	0	2	0
1	0	3	0	1	0	1	0
2	0	3	0	2	0	0	0
3	0	3	0	2	0	0	0
4	0	3	0	2	0	0	0

STATES				OPTIMAL POLICY			
5	0	3	0	2	0	0	0
0	1	3	0	0	1	1	0
1	1	3	0	1	1	0	0
2	1	3	0	2	0	0	0
3	1	3	0	2	0	0	0
4	1	3	0	2	0	0	0
0	2	3	0	0	2	0	0
1	2	3	0	1	1	0	0
2	2	3	0	2	0	0	0
3	2	3	0	2	0	0	0
0	3	3	0	0	2	0	0
1	3	3	0	1	1	0	0
2	3	3	0	2	0	0	0
0	4	3	0	0	2	0	0
1	4	3	0	1	1	0	0
0	5	3	0	0	2	0	0
0	0	4	0	0	0	2	0
1	0	4	0	1	0	1	0
2	0	4	0	2	0	0	0
3	0	4	0	2	0	0	0
4	0	4	0	2	0	0	0
0	1	4	0	0	1	1	0
1	1	4	0	1	1	0	0
2	1	4	0	2	0	0	0
3	1	4	0	2	0	0	0
0	2	4	0	0	2	0	0
1	2	4	0	1	1	0	0
2	2	4	0	2	0	0	0
0	3	4	0	0	2	0	0
1	3	4	0	1	1	0	0
0	4	4	0	0	2	0	0
0	0	5	0	0	0	2	0
1	0	5	0	1	0	1	0
2	0	5	0	2	0	0	0
3	0	5	0	2	0	0	0
0	1	5	0	0	1	1	0
1	1	5	0	1	1	0	0
2	1	5	0	2	0	0	0
0	2	5	0	0	2	0	0
1	2	5	0	1	1	0	0
0	3	5	0	0	2	0	0
0	0	6	0	0	0	2	0
1	0	6	0	1	0	1	0
2	0	6	0	2	0	0	0
0	1	6	0	0	1	1	0
1	1	6	0	1	1	0	0
0	2	6	0	0	2	0	0
0	0	7	0	0	0	2	0
1	0	7	0	1	0	1	0
0	1	7	0	0	1	1	0
0	0	8	0	0	0	2	0
0	0	0	1	0	0	0	1
1	0	0	1	1	0	0	1
2	0	0	1	2	0	0	0
3	0	0	1	2	0	0	0
4	0	0	1	2	0	0	0

STATES				OPTIMAL POLICY			
5	0	0	1	2	0	0	0
6	0	0	1	2	0	0	0
7	0	0	1	2	0	0	0
0	1	0	1	0	1	0	1
1	1	0	1	1	1	0	0
2	1	0	1	2	0	0	0
3	1	0	1	2	0	0	0
4	1	0	1	2	0	0	0
5	1	0	1	2	0	0	0
6	1	0	1	2	0	0	0
0	2	0	1	0	2	0	0
1	2	0	1	1	1	0	0
2	2	0	1	2	0	0	0
3	2	0	1	2	0	0	0
4	2	0	1	2	0	0	0
5	2	0	1	2	0	0	0
0	3	0	1	0	2	0	0
1	3	0	1	1	1	0	0
2	3	0	1	2	0	0	0
3	3	0	1	2	0	0	0
4	3	0	1	2	0	0	0
0	4	0	1	0	2	0	0
1	4	0	1	1	1	0	0
2	4	0	1	2	0	0	0
3	4	0	1	2	0	0	0
0	5	0	1	0	2	0	0
1	5	0	1	1	1	0	0
2	5	0	1	2	0	0	0
0	6	0	1	0	2	0	0
1	6	0	1	1	1	0	0
0	7	0	1	0	2	0	0
0	0	1	1	0	0	1	1
1	0	1	1	1	0	1	0
2	0	1	1	2	0	0	0
3	0	1	1	2	0	0	0
4	0	1	1	2	0	0	0
5	0	1	1	2	0	0	0
6	0	1	1	2	0	0	0
0	1	1	1	0	1	1	0
1	1	1	1	1	1	0	0
2	1	1	1	2	0	0	0
3	1	1	1	2	0	0	0
4	1	1	1	2	0	0	0
5	1	1	1	2	0	0	0
0	2	1	1	0	2	0	0
1	2	1	1	1	1	0	0
2	2	1	1	2	0	0	0
3	2	1	1	2	0	0	0
4	2	1	1	2	0	0	0
0	3	1	1	0	2	0	0
1	3	1	1	1	1	0	0
2	3	1	1	2	0	0	0
3	3	1	1	2	0	0	0
0	4	1	1	0	2	0	0
1	4	1	1	1	1	0	0

STATES				OPTIMAL POLICY			
2	4	1	1	2	0	0	0
0	5	1	1	0	2	0	0
1	5	1	1	1	1	0	0
0	6	1	1	0	2	0	0
0	0	2	1	0	0	2	0
1	0	2	1	1	0	1	0
2	0	2	1	2	0	0	0
3	0	2	1	2	0	0	0
4	0	2	1	2	0	0	0
5	0	2	1	2	0	0	0
0	1	2	1	0	1	1	0
1	1	2	1	1	1	0	0
2	1	2	1	2	0	0	0
3	1	2	1	2	0	0	0
4	1	2	1	2	0	0	0
0	2	2	1	0	2	0	0
1	2	2	1	1	1	0	0
2	2	2	1	2	0	0	0
3	2	2	1	2	0	0	0
0	3	2	1	0	2	0	0
1	3	2	1	1	1	0	0
2	3	2	1	2	0	0	0
0	4	2	1	0	2	0	0
1	4	2	1	1	1	0	0
0	5	2	1	0	2	0	0
0	0	3	1	0	0	2	0
1	0	3	1	1	0	1	0
2	0	3	1	2	0	0	0
3	0	3	1	2	0	0	0
4	0	3	1	2	0	0	0
0	1	3	1	0	1	1	0
1	1	3	1	1	1	0	0
2	1	3	1	2	0	0	0
3	1	3	1	2	0	0	0
0	2	3	1	0	2	0	0
1	2	3	1	1	1	0	0
2	2	3	1	2	0	0	0
0	3	3	1	0	2	0	0
1	3	3	1	1	1	0	0
0	4	3	1	0	2	0	0
0	0	4	1	0	0	2	0
1	0	4	1	1	0	1	0
2	0	4	1	2	0	0	0
3	0	4	1	2	0	0	0
0	1	4	1	0	1	1	0
1	1	4	1	1	1	0	0
2	1	4	1	2	0	0	0
0	2	4	1	0	2	0	0
1	2	4	1	1	1	0	0
0	3	4	1	0	2	0	0
0	0	5	1	0	0	2	0
1	0	5	1	1	0	1	0
2	0	5	1	2	0	0	0
0	1	5	1	0	1	1	0
1	1	5	1	1	1	0	0

STATES				OPTIMAL POLICY			
0	2	5	1	0	2	0	0
0	0	6	1	0	0	2	0
1	0	6	1	1	0	1	0
0	1	6	1	0	1	1	0
0	0	7	1	0	0	2	0
0	0	0	2	0	0	0	2
1	0	0	2	1	0	0	1
2	0	0	2	2	0	0	0
3	0	0	2	2	0	0	0
4	0	0	2	2	0	0	0
5	0	0	2	2	0	0	0
6	0	0	2	2	0	0	0
0	1	0	2	0	1	0	1
1	1	0	2	1	1	0	0
2	1	0	2	2	0	0	0
3	1	0	2	2	0	0	0
4	1	0	2	2	0	0	0
5	1	0	2	2	0	0	0
0	2	0	2	0	2	0	0
1	2	0	2	1	1	0	0
2	2	0	2	2	0	0	0
3	2	0	2	2	0	0	0
4	2	0	2	2	0	0	0
0	3	0	2	0	2	0	0
1	3	0	2	1	1	0	0
2	3	0	2	2	0	0	0
3	3	0	2	2	0	0	0
0	4	0	2	0	2	0	0
1	4	0	2	1	1	0	0
2	4	0	2	2	0	0	0
0	5	0	2	0	2	0	0
1	5	0	2	1	1	0	0
0	6	0	2	0	2	0	0
0	0	1	2	0	0	1	1
1	0	1	2	1	0	1	0
2	0	1	2	2	0	0	0
3	0	1	2	2	0	0	0
4	0	1	2	2	0	0	0
5	0	1	2	2	0	0	0
0	1	1	2	0	1	1	0
1	1	1	2	1	1	0	0
2	1	1	2	2	0	0	0
3	1	1	2	2	0	0	0
4	1	1	2	2	0	0	0
0	2	1	2	0	2	0	0
1	2	1	2	1	1	0	0
2	2	1	2	2	0	0	0
3	2	1	2	2	0	0	0
0	3	1	2	0	2	0	0
1	3	1	2	1	1	0	0
2	3	1	2	2	0	0	0
0	4	1	2	0	2	0	0
1	4	1	2	1	1	0	0
0	5	1	2	0	2	0	0
0	0	2	2	0	0	2	0
1	0	2	2	1	0	1	0
2	0	2	2	2	0	0	0
3	0	2	2	2	0	0	0

STATES				OPTIMAL POLICY			
4	0	2	2	2	0	0	0
0	1	2	2	0	1	1	0
1	1	2	2	1	1	0	0
2	1	2	2	2	0	0	0
3	1	2	2	2	0	0	0
0	2	2	2	0	2	0	0
1	2	2	2	1	1	0	0
2	2	2	2	2	0	0	0
0	3	2	2	0	2	0	0
1	3	2	2	1	1	0	0
0	4	2	2	0	2	0	0
0	0	3	2	0	0	2	0
1	0	3	2	1	0	1	0
2	0	3	2	2	0	0	0
3	0	3	2	2	0	0	0
0	1	3	2	0	1	1	0
1	1	3	2	1	1	0	0
2	1	3	2	2	0	0	0
0	2	3	2	0	2	0	0
1	2	3	2	1	1	0	0
0	3	3	2	0	2	0	0
0	0	4	2	0	0	2	0
1	0	4	2	1	0	1	0
2	0	4	2	2	0	0	0
0	1	4	2	0	1	1	0
1	1	4	2	1	1	0	0
0	2	4	2	0	2	0	0
0	0	5	2	0	0	2	0
1	0	5	2	1	0	1	0
0	1	5	2	0	1	1	0
0	0	6	2	0	0	2	0
0	0	0	3	0	0	0	2
1	0	0	3	1	0	0	1
2	0	0	3	2	0	0	0
3	0	0	3	2	0	0	0
4	0	0	3	2	0	0	0
5	0	0	3	2	0	0	0
0	1	0	3	0	1	0	1
1	1	0	3	1	1	0	0
2	1	0	3	2	0	0	0
3	1	0	3	2	0	0	0
4	1	0	3	2	0	0	0
0	2	0	3	0	2	0	0
1	2	0	3	1	1	0	0
2	2	0	3	2	0	0	0
3	2	0	3	2	0	0	0
0	3	0	3	0	2	0	0
1	3	0	3	1	1	0	0
2	3	0	3	2	0	0	0
0	4	0	3	0	2	0	0
1	4	0	3	1	1	0	0
0	5	0	3	0	2	0	0
0	0	1	3	0	0	1	1
1	0	1	3	1	0	1	0
2	0	1	3	2	0	0	0
3	0	1	3	2	0	0	0
4	0	1	3	2	0	0	0
0	1	1	3	0	1	1	0

STATES				OPTIMAL POLICY			
1	1	1	3	1	1	0	0
2	1	1	3	2	0	0	0
3	1	1	3	2	0	0	0
0	2	1	3	0	2	0	0
1	2	1	3	1	1	0	0
2	2	1	3	2	0	0	0
0	3	1	3	0	2	0	0
1	3	1	3	1	1	0	0
0	4	1	3	0	2	0	0
0	0	2	3	0	0	2	0
1	0	2	3	1	0	1	0
2	0	2	3	2	0	0	0
3	0	2	3	2	0	0	0
0	1	2	3	0	1	1	0
1	1	2	3	1	1	0	0
2	1	2	3	2	0	0	0
0	2	2	3	0	2	0	0
1	2	2	3	1	1	0	0
0	3	2	3	0	2	0	0
0	0	3	3	0	0	2	0
1	0	3	3	1	0	1	0
2	0	3	3	2	0	0	0
0	1	3	3	0	1	1	0
1	1	3	3	1	1	0	0
0	2	3	3	0	2	0	0
0	0	4	3	0	0	2	0
1	0	4	3	1	0	1	0
0	1	4	3	0	1	1	0
0	0	5	3	0	0	2	0
0	0	0	4	0	0	0	2
1	0	0	4	1	0	0	1
2	0	0	4	2	0	0	0
3	0	0	4	2	0	0	0
4	0	0	4	2	0	0	0
0	1	0	4	0	1	0	1
1	1	0	4	1	1	0	0
2	1	0	4	2	0	0	0
3	1	0	4	2	0	0	0
0	2	0	4	0	2	0	0
1	2	0	4	1	1	0	0
2	2	0	4	2	0	0	0
0	3	0	4	0	2	0	0
1	3	0	4	1	1	0	0
0	4	0	4	0	2	0	0
0	0	1	4	0	0	1	1
1	0	1	4	1	0	1	0
2	0	1	4	2	0	0	0
3	0	1	4	2	0	0	0
0	1	1	4	0	1	1	0
1	1	1	4	1	1	0	0

STATES				OPTIMAL POLICY			
2	1	1	4	2	0	0	0
0	2	1	4	0	2	0	0
1	2	1	4	1	1	0	0
0	3	1	4	0	2	0	0
0	0	2	4	0	0	2	0
1	0	2	4	1	0	1	0
2	0	2	4	2	0	0	0
0	1	2	4	0	1	1	0
1	1	2	4	1	1	0	0
0	2	2	4	0	2	0	0
0	0	3	4	0	0	2	0
1	0	3	4	1	0	1	0
0	1	3	4	0	1	1	0
0	0	4	4	0	0	2	0
0	0	0	5	0	0	0	2
1	0	0	5	1	0	0	1
2	0	0	5	2	0	0	0
3	0	0	5	2	0	0	0
0	1	0	5	0	1	0	1
1	1	0	5	1	1	0	0
2	1	0	5	2	0	0	0
0	2	0	5	0	2	0	0
1	2	0	5	1	1	0	0
0	3	0	5	0	2	0	0
0	0	1	5	0	0	1	1
1	0	1	5	1	0	1	0
2	0	1	5	2	0	0	0
0	1	1	5	0	1	1	0
1	1	1	5	1	1	0	0
0	2	1	5	0	2	0	0
0	0	2	5	0	0	2	0
1	0	2	5	1	0	1	0
0	1	2	5	0	1	1	0
0	0	3	5	0	0	2	0
0	0	0	6	0	0	0	2
1	0	0	6	1	0	0	1
2	0	0	6	2	0	0	0
0	1	0	6	0	1	0	1
1	1	0	6	1	1	0	0
0	2	0	6	0	2	0	0
0	0	1	6	0	0	1	1
1	0	1	6	1	0	1	0
0	1	1	6	0	1	1	0
0	0	2	6	0	0	2	0
0	0	0	7	0	0	0	2
1	0	0	7	1	0	0	1
0	1	0	7	0	1	0	1
0	0	1	7	0	0	1	1
0	0	0	8	0	0	0	2

Appendix C

Optimal policy of Model 1b

In this appendix, the optimal policy for Model 1b when $\lambda_6 = 0, \lambda_{24} = 0.817, s = 1$ using CC_1 (of Model 1b) is shown in the table below.

Table C.1: Optimal policy Model 1b when $\lambda_6 = 0, \lambda_{24} = 0.817, s = 1$ using CC_1

STATES				OPTIMAL POLICY				
0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0
2	0	0	0	1	0	0	0	0
3	0	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0	0
0	1	0	0	0	1	0	0	0
1	1	0	0	1	0	0	0	0
2	1	0	0	1	0	0	0	0
3	1	0	0	1	0	0	0	0
0	2	0	0	0	1	0	0	0
1	2	0	0	1	0	0	0	0
2	2	0	0	1	0	0	0	0
0	3	0	0	0	1	0	0	0
1	3	0	0	0	1	0	0	0
0	4	0	0	0	1	0	0	0
0	0	1	0	0	0	1	0	0
1	0	1	0	1	0	0	0	0
2	0	1	0	1	0	0	0	0
3	0	1	0	1	0	0	0	0
0	1	1	0	0	1	0	0	0
1	1	1	0	1	0	0	0	0
2	1	1	0	1	0	0	0	0
0	2	1	0	0	1	0	0	0
1	2	1	0	1	0	0	0	0
0	3	1	0	0	1	0	0	0
0	0	2	0	0	0	1	0	0
1	0	2	0	1	0	0	0	0
2	0	2	0	1	0	0	0	0
0	1	2	0	0	1	0	0	0
1	1	2	0	1	0	0	0	0
0	2	2	0	0	1	0	0	0
0	0	3	0	0	0	1	0	0
1	0	3	0	1	0	0	0	0
0	1	3	0	0	1	0	0	0
0	0	4	0	0	0	1	0	0
0	0	0	1	0	0	0	1	0
1	0	0	1	1	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	1	1	0	0	0	0
0	1	0	1	0	1	0	0	0
1	1	0	1	1	0	0	0	0
2	1	0	1	1	0	0	0	0
0	2	0	1	0	1	0	0	0
1	2	0	1	1	0	0	0	0
0	3	0	1	0	1	0	0	0
0	0	1	1	0	0	1	0	0
1	0	1	1	1	0	0	0	0
2	0	1	1	1	0	0	0	0
0	1	1	1	0	0	0	1	0
1	1	1	1	1	0	0	0	0
0	2	1	1	0	1	0	0	0
0	0	2	1	0	0	0	1	0
1	0	2	1	1	0	0	0	0
0	1	2	1	0	1	0	0	0
0	0	3	1	0	0	1	0	0
0	0	0	2	0	0	0	1	0

STATES				OPTIMAL POLICY				
1	0	0	2	1	0	0	0	0
2	0	0	2	1	0	0	0	0
0	1	0	2	0	1	0	0	0
1	1	0	2	1	0	0	0	0
0	2	0	2	0	1	0	0	0
0	0	1	2	0	0	1	0	0
1	0	1	2	1	0	0	0	0

STATES				OPTIMAL POLICY				
0	1	1	2	0	1	0	0	0
0	0	2	2	0	0	1	0	0
0	0	0	3	0	0	0	1	0
1	0	0	3	1	0	0	0	0
0	1	0	3	0	1	0	0	0
0	0	1	3	0	0	1	0	0
0	0	0	4	0	0	0	1	0

Appendix D

Optimal policy of Model 2

In this chapter, we present the optimal policy for Model 2 when $\lambda_6 = 0.45, \lambda_{24} = 0.5$, $s = 1$ and $c_e > \hat{c}_e$ in the following table.

Table D.1: Optimal policy Model 2 when $\lambda_6 = 0.45, \lambda_{24} = 0.5$, $s = 1$ and $c_e > \hat{c}_e$

STATES						OPTIMAL POLICY				
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	0	0
1	0	0	0	1	0	0	1	0	0	0
0	1	0	0	1	0	1	0	0	0	0
1	1	0	0	1	0	0	0	1	0	0
0	0	1	0	1	0	1	0	0	0	0
1	0	1	0	1	0	0	1	0	0	0
0	1	1	0	1	0	1	0	0	0	0
1	1	1	0	1	0	0	0	0	1	0
1	0	0	1	1	0	0	1	0	0	0
0	1	0	1	1	0	1	0	0	0	0
1	1	0	1	1	0	0	0	1	0	0
0	0	1	1	1	0	1	0	0	0	0
1	0	1	1	1	0	0	1	0	0	0
0	1	1	1	1	0	1	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0

STATES						OPTIMAL POLICY				
1	0	0	0	0	1	0	0	0	1	0
0	1	0	0	0	1	1	0	0	0	0
1	1	0	0	0	1	0	0	0	1	0
0	0	1	0	0	1	1	0	0	0	0
1	0	1	0	0	1	0	0	0	1	0
0	1	1	0	0	1	1	0	0	0	0
1	1	1	0	0	1	0	0	0	1	0
0	0	0	1	0	1	1	0	0	0	0
1	0	0	1	0	1	0	0	1	0	0
0	1	0	1	0	1	1	0	0	0	0
1	1	0	1	0	1	0	0	1	0	0
0	0	1	1	0	1	1	0	0	0	0
1	0	1	1	0	1	0	1	0	0	0
0	1	1	1	0	1	1	0	0	0	0
0	0	0	0	1	1	1	0	0	1	0
1	0	0	0	1	1	0	1	0	1	0
0	1	0	0	1	1	1	0	0	1	0
1	1	0	0	1	1	0	0	1	1	0
0	0	1	0	1	1	1	0	0	1	0
1	0	1	0	1	1	0	1	0	1	0
0	1	1	0	1	1	1	0	0	1	0
0	0	0	1	1	1	1	0	1	0	0
1	0	0	1	1	1	0	1	1	0	0
0	1	0	1	1	1	1	0	1	0	0
0	0	1	1	1	1	1	1	0	0	0
0	0	0	0	0	2	1	0	0	1	0
1	0	0	0	0	2	0	0	1	1	0
0	1	0	0	0	2	1	0	0	1	0
1	1	0	0	0	2	0	0	1	1	0
0	0	1	0	0	2	1	0	0	1	0
1	0	1	0	0	2	0	1	0	1	0
0	1	1	0	0	2	1	0	0	1	0
0	0	0	1	0	2	1	0	1	0	0
1	0	0	1	0	2	0	1	1	0	0
0	1	0	1	0	2	1	0	1	0	0
0	0	1	1	0	2	1	1	0	0	0
0	0	0	0	1	2	1	0	1	1	0
1	0	0	0	1	2	0	1	1	1	0
0	1	0	0	1	2	1	0	1	1	0
0	0	1	0	1	2	1	1	0	1	0
0	0	0	1	1	2	1	1	1	0	0
0	0	0	0	0	3	1	0	1	1	0
1	0	0	0	0	3	0	1	1	1	0
0	1	0	0	0	3	1	0	1	1	0
0	0	1	0	0	3	1	1	0	1	0
0	0	0	1	0	3	1	1	1	0	0
0	0	0	0	1	3	1	1	1	1	0
0	0	0	0	0	4	1	1	1	1	0
0	0	0	1	1	0	1	0	0	0	0
1	1	0	0	1	1	0	0	1	1	0
0	0	1	0	1	1	1	0	0	1	0
1	0	1	0	1	1	0	1	0	1	0
0	1	1	0	1	1	1	0	0	1	0
0	0	0	1	1	1	1	0	1	0	0
1	0	0	1	1	1	0	1	1	0	0
0	1	0	1	1	1	1	0	1	0	0
0	0	1	1	1	1	1	1	0	0	0
0	0	0	0	0	2	1	0	0	1	0

STATES						OPTIMAL POLICY				
1	0	0	0	0	2	0	0	1	1	0
0	1	0	0	0	2	1	0	0	1	0
1	1	0	0	0	2	0	0	1	1	0
0	0	1	0	0	2	1	0	0	1	0
1	0	1	0	0	2	0	1	0	1	0
0	1	1	0	0	2	1	0	0	1	0
0	0	0	1	0	2	1	0	1	0	0
1	0	0	1	0	2	0	1	1	0	0
0	1	0	1	0	2	1	0	1	0	0
0	0	1	1	0	2	1	1	0	0	0
0	0	0	0	1	2	1	0	1	1	0
1	0	0	0	1	2	0	1	1	1	0
0	1	0	0	1	2	1	0	1	1	0
0	0	1	0	1	2	1	1	0	1	0
0	0	0	1	1	2	1	1	1	0	0
0	0	0	0	0	3	1	0	1	1	0
1	0	0	0	0	3	0	1	1	1	0
0	1	0	0	0	3	1	0	1	1	0
0	0	1	0	0	3	1	1	0	1	0
0	0	0	1	0	3	1	1	1	0	0
0	0	0	0	1	3	1	1	1	1	0
0	0	0	0	0	4	1	1	1	1	0
0	0	0	1	1	0	1	0	0	0	0

Appendix E

Limiting distribution of the optimal policy

In this appendix, we present the limiting distribution resulted from using the optimal policy in each MDP model that we proposed in Chapter 4. Limiting distribution shows the stationary probability we end up in each state when we schedule the surgeries according to the optimal policy.

E.1 Limiting distribution Model 1

In this section, we present the limiting distribution from the optimal policy of Model 1. First, we present the top five limiting probability of the optimal policy resulting from Model 1 for some cases where the arrival rates are close to s. Consider $\lambda_6 = 0.05; \lambda_{24} = 0.05; s = 1$, using cost combination CC_1 in Table 5.1, we have five largest limiting probabilities in the following table.

State				Limiting probability
0	0	0	0	0.901
1	0	0	0	0.046
0	0	0	1	0.045
0	0	1	0	0.003
1	0	0	1	0.002

Table E.1: Top 5 limiting probability of Model 1's optimal policy for $\lambda_6 = 0.05; \lambda_{24} = 0.05; s = 1$

Recall that in Model 1 rejecting 6-hour patient arrivals happens when the first element of the state is equal to s. Meanwhile, the 24-hour patient arrivals are rejected when the total number of patients in the system is $4s$. In this case, the probabilities of rejecting 6-hour and 24-hour patients are 0.049 and 2.055×10^{-6}., respectively.

Next, we consider the cases where arrival rates are close to s. For $\lambda_6 = 0; \lambda_{24} = 0.817; s = 1$ using CC_1 the highest five limiting probability is in the following table.

State				Limiting probability
0	0	0	0	0.901
1	0	0	0	0.046
0	0	0	1	0.045
0	0	1	0	0.003
1	0	0	1	0.002

Table E.2: Top 5 limiting probability of Model 1's optimal policy for $\lambda_6 = 0; \lambda_{24} = 0.817; s = 1$

The probabilities of rejecting 6-hour and 24-hour patients are 0.378 and 0.053, respectively.

The five highest limiting probability for $\lambda_6 = 0.45; \lambda_{24} = 0.5; s = 1$ using CC_1 is in the following table.

State				Limiting probability
0	0	0	0	0.214
1	0	0	0	0.144
0	0	0	1	0.107
1	0	0	1	0.072
0	0	1	0	0.055

Table E.3: Top 5 limiting probability of Model 1's optimal policy for $\lambda_6 = 0.45; \lambda_{24} = 0.5; s = 1$

The probabilities of rejecting 6-hour and 24-hour patients are 0.423 and 0.064, respectively.

Another case where the total arrival rates is really close to s is $\lambda_6 = 0.54; \lambda_{24} = 1.45; s = 2$. Using CC_1, the following table shows the five highest limiting probability for that case.

State				Limiting probability
0	0	0	1	0.056
0	0	0	2	0.043
0	0	0	0	0.036
1	0	0	1	0.026
0	0	1	1	0.025

Table E.4: Top 5 limiting probability of Model 1's optimal policy for $\lambda_6 = 0.54; \lambda_{24} = 1.54; s = 2$

For this case, the probabilities of rejecting 6-hour and 24-hour patients are 0.228 and 0.088, respectively. Next, we present the limiting distribution resulted from the optimal policy of Model 1b.

E.2 Limiting distribution Model 1b

We provide the limiting probability of the optimal policy in Model 1b for some cases where the arrival rates are both close to s and far from s. First, for $\lambda_6 = 0.05; \lambda_{24} = 0.2$, using cost combination CC_1 in Table 5.5, we have five largest limiting probabilities in the following table.

State				Limiting probability
0	0	0	0	0.777
0	0	0	1	0.155
1	0	0	0	0.039
0	0	0	2	0.016
1	0	0	1	0.008

Table E.5: Top 5 limiting probability of Model 1b's optimal policy for $\lambda_6 = 0.05; \lambda_{24} = 0.2; s = 2$

From the table above, we can see that using the optimal policy from Model 1b, around 78% of the time, we have no surgeries waiting. It is a large probability of being in the state that is empty. Hence, in this case, the optimal policy provides a good scheduling considering the arrival rates and the dedicated capacity, as no patients are piled up most of the time.

Another case where the arrival rates are close to s, we look at the limiting probability for $\lambda_6 = 0.05; \lambda_{24} = 0.05; s = 1$ using CC_1 in the following table.

State				Limiting probability
0	0	0	0	0.901
1	0	0	0	0.046
0	0	0	1	0.045
0	0	1	0	0.003
1	0	0	1	0.002

Table E.6: Top 5 limiting probability of Model 1b's optimal policy for $\lambda_6 = 0.05; \lambda_{24} = 0.05; s = 1$

From the limiting probability, we also obtain that the probability of 6-hour patients being rejected is 0.049, while the 24-hour patients are rejected with probability of 2.055×10^{-6}.

Next, using CC_1 and CC_3 the following table shows the limiting probability when $\lambda_6 = 0.54; \lambda_{24} = 1.45$ and $s = 2$. We can see that for CC_3, where the cost of deferring 24-hour patients is zero, the probabilities of ending up in the states where the system is not full is higher than those of CC_1. For example, using cost combination CC_3, there is around 14% chance of being in empty state, while for CC_1, the chance of being in this state is only 4.7%. This results match with intuition as in CC_3, the 24-hour patients that can not be treated in the dedicated capacity are deferred to another resource. Meanwhile, in CC_1 less 24-hour patients are deferred due to the non-zero cost of deferring patients.

State				Limiting probability
0	0	0	1	0.198
0	0	0	2	0.144
0	0	0	0	0.137
1	0	0	1	0.107
1	0	0	2	0.078

Table E.7: Limiting probability: Case 3;CC_3

State				Limiting probability
0	0	0	1	0.069
0	0	0	2	0.050
0	0	0	0	0.047
1	0	0	1	0.038
0	0	1	1	0.030

Table E.8: Limiting probability: Case 3;CC_1

Table E.9: Top 5 limiting probability of Model 1b's optimal policy for $\lambda_6 = 0.54; \lambda_{24} = 1.45$ and $s = 2$

In cost combinations CC_1 and CC_2 the deferring costs are non-zero. For these cases, the five states with the largest limiting probabilities are the same (Table E.12). Similar to the previous tables, the higher deferring cost results in lower limiting probabilities. This is because in CC_2 less 24-hour patients are deferred due to the higher cost. Hence, the probabilities of being in the more crowded states are higher than those in CC_1 case.

State				Limiting probability
0	0	0	1	0.069
0	0	0	2	0.050
0	0	0	0	0.047
1	0	0	1	0.038
0	0	1	1	0.030

Table E.10: Limiting probability: Case 3;CC_1

State				Limiting probability
0	0	0	1	0.060
0	0	0	2	0.043
0	0	0	0	0.041
1	0	0	1	0.034
0	0	1	1	0.027

Table E.11: Limiting probability: Case 3;CC_2

Table E.12: Top 5 limiting probability of Model 1b's optimal policy for $\lambda_6 = 0.54; \lambda_{24} = 1.45$ and $s = 2$

Using CC_1, 0.138 of the time 6-hour patients are rejected, while 0.016 of the time 24-hour patients are rejected.

E.3 Limiting distribution Model 2

First, we look at the limiting probability where the arrival rates are far from s using cost combination where $c_e > \hat{c}_e$. For $\lambda_6 = 0.05$ and $\lambda_{24} = 0.05$, the optimal policy which yield in five largest limiting probabilities given in the following table.

State						Limiting probability
0	0	0	0	0	0	0.895
0	0	0	0	1	0	0.046
0	0	0	0	0	1	0.045
1	0	0	0	0	0	0.003
0	1	0	0	0	0	0.003

Table E.13: Top 5 limiting probability of Model 2's optimal policy for $\lambda_6 = 0.05; \lambda_{24} = 0.05$ and $s = 1$

We reject 6-hour arrivals when dedicated capacity s is fully used in the next shift. From this limiting probability, we also obtain that with probability of 0.052 the 6-hour patients are rejected, while the 24-hour patients are rejected with probability of 2.421×10^{-6}.

Next, for $\lambda_6 = 0; \lambda_{24} = 0.05$ and $s = 1$, where $c_e > \hat{c}_e$ results in optimal policy with the top five limiting probability in the following table.

State						Limiting probability
0	0	0	0	0	0	0.950
0	0	0	0	0	1	0.048
1	0	0	0	0	0	0.001
0	0	0	0	0	2	0.001
1	0	0	0	0	1	0.00006≈0

Table E.14: Top 5 limiting probability of Model 2's optimal policy for $\lambda_6 = 0; \lambda_{24} = 0.05$ and $s = 1$

For the case above the probability of rejecting 6-hour patients is 0.001. Meanwhile, we reject the 24-hour patients when the system is full, i.e., there are $4s$ patients in the system (both the scheduled and the new admitted ones). The probability of rejecting 24-hour patients for the case above is 3.183×10^{-7}. From the two cases above, we can see that when the arrival rates and s have larger gaps, the possibility of staying in an empty state is high, i.e., more than 85%.

Afterwards, we look at the limiting probability for the case where the arrival rates are close to s. For $\lambda_6 = 0.45$; $\lambda_{24} = 0.5$ and $s = 1$, the optimal policy which yield in five largest limiting probabilities given in Table E.15.

State						Limiting probability
0	0	0	0	0	0	0.169
0	0	0	0	1	0	0.096
1	0	0	0	0	0	0.087
0	0	0	0	0	1	0.084
1	1	0	0	0	0	0.051

Table E.15: Top 5 limiting probability of Model 2's optimal policy for $\lambda_6 = 0.45; \lambda_{24} = 0.5$ and $s = 1$

For the case above, the limiting probabilities of rejecting 6-hour and 24-hour patients are 0.315 and 0.058, respectively. Next, the optimal policy of Model 1b for the case where $\lambda_6 = 0; \lambda_{24} = 0.817$ and $s = 1$ result in the five highest limiting distribution given in the table below.

State						Limiting probability
0	0	0	0	0	0	0.224
0	0	0	0	0	1	0.183
1	0	0	0	0	0	0.100
1	0	0	0	0	1	0.082
0	0	0	0	0	2	0.075

Table E.16: Top 5 limiting probability of Model 2's optimal policy for $\lambda_6 = 0; \lambda_{24} = 0.817$ and $s = 1$

Next, we look at the limiting probability for $\lambda_6 = 0.45; \lambda_{24} = 1.54 : s = 2$ both when $c_e = \hat{c}_e$ and $c_e > \hat{c}_e$ in Table E.17 and E.18, respectively.

State						Limiting probability
2	1	1	0	0	1	0.034
0	0	0	0	0	1	0.031
1	1	1	0	0	1	0.028
2	1	1	0	0	2	0.025
1	1	0	0	0	1	0.024

Table E.17: Top 5 limiting probability of Model 2's optimal policy for $\lambda_6 = 0.45; \lambda_{24} = 1.54$ and $s = 2$ when $c_e = \hat{c}_e$

State						Limiting probability
2	2	2	0	0	2	0.027
0	0	0	0	0	1	0.026
2	2	2	0	0	1	0.022
1	1	1	0	0	1	0.021
0	0	0	0	0	2	0.019

Table E.18: Top 5 limiting probability of Model 2's optimal policy for $\lambda_6 = 0.45; \lambda_{24} = 1.54$ and $s = 2$ when $c_e > \hat{c}_e$

Recall that the 6-hour patient arrivals are rejected when the capacity in the next shift is fully used, which in this case is when the number of patients scheduled next shift is 2. Using the limiting probabilities of the last case where $c_e > \hat{c}_e$ and $c_e = \hat{c}_e$, the probabilities of 6-hour patients being rejected are 0.233 and 0.241, respectively. Meanwhile, the 24-hour patients are rejected with probabilities of 0.064 and 0.045, for $c_e > \hat{c}_e$ and $c_e = \hat{c}_e$, respectively.