

DOKUZ EYLUL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**ANALYSIS OF APPOINTMENT SCHEDULING
SYSTEMS IN DEPARTMENT OF ORTHOPEDICS AND
TRAUMATOLOGY IN DEU FACULTY OF MEDICINE**

by
Sezgi ÖZEN

July, 2005
İZMİR

ANALYSIS OF APPOINTMENT SCHEDULING SYSTEMS IN DEPARTMENT OF ORTHOPEDICS AND TRAUMATOLOGY IN DEU FACULTY OF MEDICINE

**A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science in
Industrial Engineering, Industrial Engineering Program**

**by
Sezgi ÖZEN**

**July, 2005
İZMİR**

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**Analysis Of Appointment Scheduling Systems In Department Of Orthopedics And Traumatology In DEU Faculty Of Medicine**” completed by **Sezgi ÖZEN** under supervision of **Prof. Dr. İrem ÖZKARAHAN** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

.....
Prof. Dr. İrem ÖZKARAHAN

Supervisor

.....
Prof. Dr. Hasan ESKİ

(Jury Member)

.....
Prof. Dr. Alp KUT

(Jury Member)

.....
Prof.Dr. Cahit HELVACI
Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I would like to express my heart-felt thanks first and foremost to my advisor, Prof. Dr. İrem Özkarahan, for her guidance, cooperation, encouragement and advice throughout the preparation of this thesis. I would also like to express my gratitude to her for giving me the opportunity to share her experience and insights during this process in a friendly atmosphere that inspired creativity and motivation.

I would like to thank to all professors and colleagues in the Department of Industrial Engineering for their encouragement and support during my study. I am also indebted to Research Assistant Nazan Zeybekcan, as a colleague and one of my best friends, without her encouragement and motivation I truly doubt that this day would ever have arrived.

I am grateful to Dokuz Eylul University Faculty of Medicine staff, especially Jülide Çorlu, for their support during gathering the data of the study. .

I am also fortunate to have friends who have provided me with confidence and love on my desperate days. I am really grateful to all of them for their motivation and support during my study.

There is no way I can adequately express my gratitude to my family who have always provided me with endless love, security and incredible support. I am deeply indebted to my parents, Nevcihan and Aydın Özen, for their confidence, encouragement and support throughout all these years of my education, which were very crucial to my accomplishment and well-being. This thesis is dedicated to them.

Sezgi ÖZEN

**ANALYSIS OF APPOINTMENT SCHEDULING SYSTEMS
IN DEPARTMENT OF ORTHOPEDICS AND TRAUMATOLOGY
IN DEU FACULTY OF MEDICINE**

ABSTRACT

Nowadays, efficiency and effectiveness of healthcare organizations are commonly analyzed by both managers and researchers because of the increasing importance of health industry. Also, this thesis presents approaches for the optimization problems in healthcare organizations. The problems are mainly concerned with optimizing patient waiting times and doctor idle time. The healthcare organization is analyzed as two different systems. First, the problem of appointment system is solved for both dynamic and static state by using dynamic programming and branch-and-bound approaches respectively. The proposed system is a variable-sized multiple-block appointment system and optimal patient assignments for each period (block) are obtained. Then, personnel scheduling problem in the outpatient clinic is solved. To determine doctor requirements of the clinic, queuing theory is used. Data on patient arrivals is gathered for 23 weeks and by performing input analysis parameters of queuing formulas are determined. Then by using linear programming approach, optimal schedule for the doctors in the outpatient clinic is obtained.

Keywords: Appointment Scheduling, Dynamic Programming, Personnel Scheduling, Healthcare Organizations

**DOKUZ EYLÜL ÜNİVERSİTESİ TIP FAKÜLTESİ ORTOPEDİ VE
TRAVMATOLOJİ ANABİLİM DALINDA RANDEVU ÇİZELGELEME
SİSTEMLERİNİN ANALİZİ**

ÖZ

Günümüzde gitgide artan öneminden dolayı, sağlık sistemlerinin etkinliği ve verimliliği yöneticiler ve araştırmacılar tarafından daha sık incelenmeye başlanmıştır. Bu çalışmada da sağlık sistemlerindeki optimizasyon problemlerine yönelik çözüm teknikleri sunulmuştur. Sağlık sistemlerindeki mevcut problemler genel olarak hasta bekleme sürelerinin ve doktor boş kalma süresinin optimizasyonu ile ilgilidir. Bu çalışmada incelenen sağlık sistemi iki ayrı farklı şekilde ele alınmıştır. İlk olarak randevu sistemine ilişkin problem hem dinamik hem de statik durumlar için çözülmüştür. Bu çözümlerde dinamik programlama ve dal sınır algoritması yaklaşımları kullanılmıştır. Önerilen sistem değişken büyüklükte çok bloklu bir randevu sistemidir ve her bir periyot (blok) için optimum hasta atamaları hesaplanmıştır. İkinci olarak poliklinikte çalışan doktorlara ait çizelgeleme problemi çözülmüştür. Poliklinikteki doktor gereksinimi hesaplayabilmek için kuyruk teorisi kullanılmıştır. Hasta gelişlerine ait 23 haftalık veri toplanmış ve girdi analizi yapılarak kuyruk modeli formülleri uygulanmıştır. Daha sonra, doğrusal programlama yaklaşımı ile poliklinikte çalışan doktorlar için optimum çalışma saatleri hesaplanmıştır.

Anahtar Sözcükler: Randevu Sistemleri, Dinamik Programlama, Personel Çizelgeleme, Sağlık Sistemleri.

CONTENTS

	Page
THESIS EXAMINATION RESULT FORM	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ÖZ.....	v
CHAPTER 1 – INTRODUCTION	1
1.1. Motivation for the study	2
1.2. Thesis Outline.....	4
CHAPTER 2 – OVERVIEW OF APPOINTMENT SCHEDULING	6
2.1. Appointment Scheduling	6
2.2. An Appointment System	7
2.2.1 Number of Services	7
2.2.2 Number of Doctors	8
2.2.3 Number of Appointments per Clinic Session	8
2.2.4 Arrival Process.....	8
2.2.5 Service Times	10
2.2.6 Lateness and the Interruption Level of Doctors.....	10
2.2.7 Queue Discipline.....	10
2.3. Performance Measures of an Appointment System	11
2.3.1 Patient Waiting Time	11
2.3.2 Doctor Idle Time and Overtime.....	11
2.3.3 Cost Function.....	11
2.3.4 Other	12
2.4. Design of an Appointment System.....	12
2.4.1 Appointment Rules	12
2.4.2 Patient Classification	14
2.4.3 Adjustments	14

2.5. Appointment Rules	14
2.5.1 Appointment Rules Studied in the Literature	14
2.5.2 Generalized Structure for Appointment Rules.....	16
2.6. Related Research Work	16
2.6.1 Analytical Approaches to Appointment Scheduling	16
2.6.2 Simulation Approach to Appointment Scheduling.....	18
2.6.3 Case Studies in Appointment Scheduling.....	21
 CHAPTER 3 – OVERVIEW OF SOLUTION METHODOLOGIES USED FOR APPOINTMENT SCHEDULING	 24
3.1. Simulation	24
3.1.1 When Simulation Is Appropriate?	25
3.1.2 When Simulation Is Not Appropriate?	25
3.1.3 Advantages and Disadvantages of Simulation.....	26
3.2. Stochastic Linear Programming	27
3.2.1 Formulation of Stochastic Linear Program.....	28
3.3. Dynamic Programming	29
3.3.1 Characteristics of Dynamic Programming Problems.....	31
3.3.2 Deterministic Dynamic Programming.....	33
3.3.3 Probabilistic Dynamic Programming.....	34
3.3.4 Dynamic Programming in Appointment Scheduling.....	35
 CHAPTER 4 – OVERVIEW OF PERSONNEL SCHEDULING.....	 41
4.1. Personnel Scheduling	41
4.2. Literature Review	42
4.2.1 Heuristic Approaches to Personnel Scheduling.....	43
4.2.2 Self Scheduling.....	43
4.2.3 Mathematical Programming Approaches to Personnel Scheduling	44
4.3. Solution Methodology for Personnel Scheduling	45
4.3.1 Linear Programming.....	45
4.3.2 “An Integrated Nurse Scheduling Model”, Ozkarahan, Irem, 1991	46
4.4. Queuing Theory	52
4.4.1 Basic Structures of Queuing Models	53
4.4.1.1 Customer Population.....	54

4.4.1.2 Queue	55
4.4.1.3 Queue Discipline.....	55
4.4.1.4 Service Facility	55
4.4.2 Probability Distributions.....	55
4.4.2.1 Arrival Distribution.....	56
4.4.2.2 Service Time Distribution.....	56
4.4.3 Performance Measurement of Queues.....	57
4.4.3.1 The Probability for the Number of Customers in the System.....	57
4.4.3.2 The Mean Number of Customers in the System.....	57
4.4.3.3 The Mean Customer Time Spent in the System	58
4.4.3.4 Mean Number of Customers Waiting	58
4.4.3.5 The Mean Customers Waiting Time.....	58
4.4.3.6 The Server Utilization Factor.....	59
4.4.4 Basic Queuing Systems	59
4.4.5 Variety of Queuing Models	61
4.4.5.1 The Single-Channel Queuing Model	61
4.4.5.2 The Multiple-Channel Queuing Model.....	64
4.4.5.3 The Single-Channel Queuing Model for a Finite Queue	66
4.4.5.4 The Single-Channel Queuing Model with Constant Service Times	67
4.4.5.5 The Single-Channel Queuing Model for a Limited Population.....	68

CHAPTER 5 – APPLICATION IN DOKUZ EYLUL UNIVERSITY

DEPARTMENT OF ORTHOPEDICS AND TRAUMATOLOGY	70
5.1. Introduction	70
5.2. Appointment Scheduling Problem	72
5.2.1 Model Parameters	75
5.2.2 Dynamic State.....	78
5.2.2.1 Alternative Model 1	79
5.2.2.2 Alternative Model 2	89
5.2.3 Static State	92
5.2.3.1 Static Version of Alternative Model 1	93
5.2.3.2 Static Version of Alternative Model 2	106
5.3. Personnel Scheduling Problem	110
5.3.1 Queuing Theory	111

5.3.2 Input Analysis	111
5.3.2.1 Steps of Input Analysis	111
5.3.2.2 Patient Arrivals	112
5.3.2.3 Fitting Input Distributions via Input Analyzer of Arena 5.0.....	118
5.3.2.4 Service Times.....	119
5.3.3 Queuing System.....	119
5.3.3.1 Monday	120
5.3.3.2 Tuesday	125
5.3.3.3 Wednesday	131
5.3.3.4 Thursday.....	136
5.3.3.5 Friday	141
5.3.4 Personnel Scheduling Models.....	148
 CHAPTER 6 – CONCLUSION	 156
6.1. Performed Research.....	156
6.2. Future Research	157
 REFERENCES.....	 159
 APPENDIX A Patient Arrival Records	 164
APPENDIX B Total Patient Arrivals	177
APPENDIX C Arrivals per Week.....	180
APPENDIX D Patient Arrivals per Day and per Period.....	183
APPENDIX E Results of Input Analysis.....	193
APPENDIX F Solution Report of the DOW Problem	194
APPENDIX G-1 Solution Report of the TOD Problem for Monday	195
APPENDIX G-2 Solution Report of the TOD Problem for Tuesday	196
APPENDIX G-3 Solution Report of the TOD Problem for Wednesday.....	197
APPENDIX G-4 Solution Report of the TOD Problem for Thursday.....	198
APPENDIX G-5 Solution Report of the TOD Problem for Friday	199
APPENDIX H LINGO.....	200

CHAPTER ONE

INTRODUCTION

As service industries become a more important business sector, the efficiency and effectiveness of running service systems are gradually receiving more attention from both academicians and practitioners. In an increasingly competitive industry where both costs and customer expectations are rising rapidly, hospital managers are under greater pressure to manage such facilities more efficiently and effectively.

Improving outpatient resource utilization significantly enhances the efficiency of healthcare organizations. Consequently, scheduling becomes extremely complicated and important. Selecting the right scheduling alternative, a healthcare organization can improve operating efficiency of outpatient resources.

The healthcare organizations may serve customers in two ways: by appointments or through outpatient clinics that serve patients. The use of appointments is increasing day by day and it is necessary to schedule appointments of doctors in order to increase efficiency of the organization. Also, patients prefer appointment systems because they have shorter and less variable waiting times in such systems. Selecting an appropriate appointment rule can contribute the organization in terms of less waiting times for patients and less idle time for the doctor.

In the outpatient clinic, there is not an appointment based system. Patients arrive in the system randomly and generally all patients enter the system within a clinic session are treated by the doctor but such organizations typically face demand that varies throughout the day. Then, an effective personnel scheduling becomes one of the primary means by which service organizations remain competitive. In such cases, management's goal is to find the best mix of hourly employees so that demand is satisfied at minimum cost.

The health care organizations in Turkey also have two main types of services. The first one is outpatient clinics that serve all patients arrive during a period. This is commonly used among the public hospitals. Since, there is no efficient personnel scheduling system in such organizations, patients usually experience long waiting times. In order to decrease this waiting time and increasing the efficiency of the service, scheduling doctors' working hours, in order to meet the variable demand, is very important.

The second service is based on appointments. Some of private health care organizations and some doctors, who are also professors, treat patients by appointments. In order to increase efficiency of the system, regardless of it is a public or private organization, it is necessary to schedule the service. It is an important issue determining which patients will be seen by the doctor and when they will be seen. It enables the doctor to schedule his / her work load while decreasing the waiting time of the patient.

1.1 Motivation for the Study

The major goal of this study is to establish approaches of optimizing measures of performance that characterize the interest of all parties. The interest of the patients is taken only as the minimization of his/her expected waiting time while the interest of the doctor is to minimize idle time. The cost of the system is assumed to be a combination of the waiting time cost and idle time cost.

The objective of this thesis is analyzing the Department of Orthopedic and Traumatology in Dokuz Eylül University. This department is selected to analyze because it serves patients by both appointment and outpatient clinic. Then, appointment system and the outpatient clinic of the department are analyzed separately.

The team of the department, except residents, treats patients on an appointment based system. The consultation time and possible variations of consultation time are not taken into consideration while giving appointments. This cause long waiting times for patients and in some situations the doctor may be idle. Therefore, the goal of appointment scheduling study in this thesis is to optimize patient waiting time and doctor idle time by determining the optimum patient assignment for each time period.

To schedule the appointments of these doctors, appointment scheduling techniques are used. The problem is determining optimum number of patients assigned to the beginning of each time period in order to minimize patient waiting time and doctor idle time. Then, the doctor will be able to know how many patients he/she will see during each time period and patients will have less waiting times. Basic performance measures of health care organizations, patient waiting time and doctor idle time, will be used to evaluate the system. Different solutions for dynamic and static versions are developed to solve the problem. First a dynamic solution is obtained but since characteristics of dynamic programming are not too efficient for real-world problems, results are converted into static state by using branch-and-bound approach. As a result of the study, doctor will be able to know, how many patients he/she sees at what time while the patient knows at what time he/she will be seen by the doctor.

On the other hand; the rest of the team, the residents, works in the outpatient clinic. It is the busiest part of the department and has a patient demand varying throughout the day and throughout the week. The busiest period of the day is, generally, the morning period and during this period patients have too long waiting times because of the lack of doctor available in the clinic. Then, the doctors who work in the clinic need to see patients without a break. This causes a decrease in the efficiency of the organization and the quality of the service. However, in the afternoon, most of the doctors are idle because there are less patients requiring for consultation. Therefore, the goal of performing personnel scheduling in this thesis is finding the optimal schedule for doctors while meeting patient demand. Since patient

demand is determined in order to keep patient waiting time at an acceptable level, both patients' and doctors' interests are satisfied.

To find the optimal work schedule for doctors, it is necessary to determine how many doctors are necessary for each period of each day in order to meet variable patient demand. Once doctor requirements are determined for each period of each day by using queuing theory approach, a new problem arises: which doctor will work on which day and at what time. At this point, it is necessary to use personnel scheduling approaches to schedule working hours of the doctors in the clinic. Then, as a result of the study the optimal schedule for doctors in the outpatient clinic is obtained in order to meet demand and keep understaffing or overstaffing levels acceptable.

1.2 Thesis Outline

The thesis document is divided into six chapters. Chapter one contains a brief description of the scheduling problem in general. It also describes the scopes of the study.

In the next chapter of the thesis, an overview of appointment scheduling is given. Basic appointment rules used in the literature and main components of an appointment scheduling are mentioned. The performance measures of an appointment system are described and the cost function is derived. Then, a literature review is given that studies about appointment scheduling are classified in three classes.

The overview of solution methodologies used for appointment scheduling is given in Chapter three. Three basic approaches are described in this chapter: simulation, stochastic linear programming and dynamic programming. Since dynamic programming approach is applied to solve the appointment scheduling problem for doctors at Dokuz Eylul University, Department of Orthopedics and Traumatology (DEU, Dept.of Ort. and Trau.), a more detailed overview is given.

Then, a key article related to the use of dynamic programming in appointment scheduling problems is described. The formulations of this paper are used and modified for solving the appointment scheduling problem in Dokuz Eylul University, Department of Orthopedics and Traumatology.

In the next chapter, an overview of personnel scheduling, used for scheduling the doctors' working hours in the outpatient clinic, is given. A brief literature review is given and then linear programming, the basic solution methodology for personnel scheduling, is described. Finally, the key article whose formulations are used and modified in personnel scheduling problem in DEU, Dept.of Ort. and Trau. is given.

In Chapter five, application of appointment scheduling and personnel scheduling approaches in DEU, Dept.of Ort. and Trau. is given. First appointment scheduling problem is solved for both dynamic and static versions. Alternative approaches are presented and results of each approach are given. Finally, optimal number of patients scheduled at the beginning of each period is determined. Then personnel scheduling problem is solved. First queuing analysis is applied to the system and doctor requirements per period and per day are determined. Then alternative models are proposed for the solution and results of each model are given. Finally, the days of the week and the time of the day each doctor will work is determined.

Then, Chapter six states the results and the conclusion of the thesis.

CHAPTER TWO

OVERVIEW OF APPOINTMENT SCHEDULING

2.1 Appointment Scheduling

In this thesis, appointment scheduling is used to schedule the appointments of professors, associate professors and assistant professors in DEU, Dept.of Ort. and Trau. In other words, work loads of doctors are determined in terms of number of patients seen during a period and number of periods during a clinic session.

Appointment systems have flourished over the last century. With widespread availability of the telephone and other forms of improved communications, the efficiency due to scheduling got more importance. For example: no longer most people simply appear at a doctor's office for routine medical problems and wait. Instead, an appointed time of arrival is scheduled, many times by phone. This is the situation today for many personal services. It is also the practice in many industrial settings such as the scheduling of cargo ships at port facilities, the scheduling of part deliveries in just-in-time systems and the scheduling of customers at military testing and training facilities.

Customers benefit greatly from such an arrangement since their waiting times are usually both smaller and less variable under a scheduling system. Servers incur costs due to the creation and maintenance of the scheduling system. On the other hand, servers benefit from requiring smaller facilities for holding queuing customers and from increased customer satisfaction. Since customers and servers, in general, have different goals, the scheduling systems that optimize their interests will, in general, differ. The interest of each customer is the minimization of the expected waiting time. The interest of the server is to minimize overtime. So; healthcare providers are under a great deal of pressure to reduce costs and improve quality of service provided.

Two of the few tangible elements of a healthcare service are the patient waiting times and waiting room congestion. A well designed appointment system is able to increase the utilization of expensive personnel and equipment based medical resources as well as reducing waiting times for patients. It is certainly known that one of the most important reasons for patients' dissatisfaction is excessive waiting times.

Consequently; the objective of appointment scheduling is to find an appointment system for which a particular measure of performance is optimized in a clinical environment.

2.2 An Appointment System

An appointment system is like a queuing system. In the literature an appointment system can be static or dynamic. In static systems, all decisions must be made prior to the beginning of the clinic session. In dynamic systems the schedule of future arrivals is revised continuously over the day based on the current state of the system. The most common system used in healthcare is the first one, static systems.

The main components of an appointment system are as follows:

2.2.1 *Number of Services*

An appointment system may be a single-stage or multi-stage system. In a single-stage system patients wait for a single service. In a multi-stage system patients may pass through various facilities (registration, examination, laboratory... in healthcare). While modeling multi-stage systems, it is necessary to know the patient flow probabilities associated with each facility.

2.2.2 *Number of Doctors*

Most studies in literature focus on single server systems. If system has more than one server then there are two statements: doctors have their own queue or a single common queue for patients that reduce patients' waiting times.

2.2.3 *Number of Appointments per Clinic Session (N)*

Some studies in the literature report that there is a positive relationship between waiting times and the number of appointments in a clinic session. The effect of N can be reduced by no-shows and variability of consultation times.

2.2.4 *Arrival Process*

In practice, the scheduling period is always bounded and frequently is of fixed length. The horizon does not restrict the end of service or the overtime point. The arrival characteristics of patients are composed of several factors:

2.2.4.1 *Arrival Discipline*

Scheduling strategies typically are classified as individual, block or mixed. Individual strategies allow each arrival to be scheduled separately, while block strategies constrain arrivals to occur only at the beginning of each block of time. Blocks need not be the same size or contain the same number of customers. Mixed strategies typically consist of a set of simultaneous arrivals at the beginning of each block with the remainder of the customers scheduled individually.

2.2.4.2 *Customer Punctuality*

It is the difference between a patient's scheduled appointment time and actual arrival time. In practice, patient earliness is more common than patient lateness but patient earliness is undesirable since it creates excessive congestion in the waiting

area. Patient lateness is modeled as an independent variable with a certain limit on maximum lateness. In the literature, some authors combined patient and doctor unpunctuality under one variable: “system earliness”.

2.2.4.3 No-Shows

A no-show may be considered as an infinitely late customer. Many studies have attempted to investigate possible variables that might affect patient attendance and some identify policies aimed at analyzing effects of no-shows. In the literature, no-show probabilities (p) range from 0.05 to 0.3. As p increases, idleness of doctor increases while patient waiting time decreases.

2.2.4.4 Walk-Ins

Walk-ins may be regular or emergent patients who require immediate medical attention and may possibly interrupt the current consultation. They must be planned for an efficient appointment system. Probability of walk-ins varies across specialties as no-show probability. In the literature, some studies proved that arrival rates of emergency patients and walk-ins differ significantly throughout the day but not from day to day. As a result it is found that Poisson distribution offers an acceptable representation of such arrivals.

2.2.4.5 Companions

These are people who accompany a patient to the clinic. Although they do not receive the service, they utilize the waiting room and they must be considered while planning the size of a clinic’s waiting room area.

2.2.5 *Service Times*

Service or consultation time can be defined as the sum of all the times a patient is demanding the doctor's attention, preventing him/her from seeing other patients. In studies it is generally assumed patients are homogeneous for scheduling purposes and use independent and identically distributed service times. In these studies a variety of service time distributions are used but mostly Erlang or exponential service times are used to make the models tractable.

2.2.6 *Lateness and Interruption Level of Doctors*

Patient waiting times are highly sensitive to this factor. If the doctor does not start the clinic on time, a delay factor builds up from the start that affects the whole clinic session. Interruption level – gap times – includes all activities during the session that may require doctor's attention, such as interactions with support staff, phone calls, writing up notes, comfort breaks... which interrupt consultation. It is likely that patients arrive early to “beat” the system or arrive late knowing that they will have to wait anyway.

2.2.7 *Queue Discipline*

Almost in all studies in the literature patients are served on a first-come, first-served (FCFS) basis. Given punctual patients, this queue discipline is identical to serving patients in the order of their appointment times. However; unpunctuality may cause changes in the actual queue discipline, as doctors would not keep idle waiting for the next appointment in the presence of other waiting patients. A clinic which deals with walk-ins, emergencies and/or second consultations needs to set a priority rule which determines the order in which these patients will be seen. In general, the first priority is given to walk-ins. In practice, it is not uncommon for patients to be called in the order of arrival even when there is an appointment system. This may destroy the whole purpose of the system and may lead to patients ignoring appointments and coming earlier than necessary (Cayırlı&Veral, 2003).

2.3 Performance Measures Of An Appointment System

There is a variety of performance criteria used in the literature to evaluate Appointment Systems. “Mean waiting time of patients”, “Mean idle time of doctor”, “Mean overtime of doctor” are the most common performance measures. By giving them relative weights, objective becomes the minimization of expected total cost of the system:

$$\text{Min}E(\text{TC})=E(W)C_p + E(I)C_d + E(O)C_o$$

where $E(W)$, $E(I)$, $E(O)$ are the expected costs of waiting time, idle time, overtime respectively and C_p , C_d , C_o are the relative weights of expected waiting cost, idle time cost, overtime cost respectively.

2.3.1 Patient Waiting Time

“True” waiting time of patients is obtained by subtracting the greater of {appointment time, arrival time} from the consultation start time. In practice waiting due to early arrival is not taken into consideration because it is voluntary.

2.3.2 Doctor Idle Time and Overtime

Idle time of a doctor is the total time during the clinic session when he is not consulting because there are no patients waiting to be seen. Overtime is calculated as the positive difference between the desired completion time of the clinic session and the actual end of service for the last patient. In general, the negative overtime value can be considered as a part of idle time.

2.3.3 Cost Function

Cost function includes different combinations of patient’s waiting time, doctor’s idle time and doctor’s over time. But in some studies patient’s flow time is used instead of patient’s waiting time.

2.3.4 Other

Some other measures used to evaluate an appointment system are doctor's productivity, mean doctor utilization, delays between requests and clinic effectiveness (Cayırlı&Veral, 2003).

2.4 Design Of An Appointment System

2.4.1 Appointment Rules

Variables of an appointment rule are:

Block Size (n_i): This is the number of patients scheduled to the i^{th} block. Patients can be called individually, in groups of constant size or in variable block sizes.

Begin Block (n_1): This is also called initial block and it is the number of patients given an identical appointment time at the start of the session.

Appointment Interval (a_i): This is the interval between two successive appointment times. It is also called "job allowance".

Appointment rules are obtained by the combination of these three rules:

i : block number;

n_i : block size for the i^{th} block

n_1 : initial block

a_i : appointment interval

N : number of patients per clinic session

2.4.1.1 Single-Block Rule:

$$n_1 = N$$

no a_i

Patients wait for excessive times and doctors never stay idle.

2.4.1.2 Individual Block / Fixed Interval Rule:

$$n_i=1 \quad \text{for all } i$$

$$a_i = \text{constant}$$

2.4.1.3 Individual Block / Fixed Interval with Initial Block:

$$n_1 > 1$$

$$n_i = 1$$

$$a_i = \text{constant}$$

The objective of this rule is to keep inventory of patients to minimize doctor's idleness.

2.4.1.4 Multiple Block / Fixed Interval Rule:

$$n_i = m > 1 \quad \text{for all } i$$

$$a_i = \text{constant}$$

This rule is suitable when mean consultation times are short.

2.4.1.5 Multiple Block / Fixed Interval Rule with an Initial Block:

$$n_1 > m$$

$$n_i = m$$

$$a_i = \text{constant}$$

2.4.1.6 Variable Block / Fixed Interval Rule:

$$n_i \text{ variable}$$

$$a_i = \text{constant}$$

2.4.1.7 Individual Block / Variable Interval Rule:

$$n_i = 1 \quad \text{for all } i$$

$$a_i = \text{variable}$$

2.4.2 Patient Classification

Patient classification is used for two purposes:

1. To sequence patients at the time of booking
2. To adjust appointment intervals based on the distinct service time characteristics of different patient classes.

Patients can be classified according to their variability of service times, type of procedure or their attendance.

2.4.3 Adjustments

During the design of an AS some other factors such as no-shows, urgent patients, walk-ins and emergencies must be taken into consideration. To adjust no-shows there are two ways used in the literature. It is possible to overbook extra patients to predetermined slots or to decrease appointment intervals. To adjust the other factors (walk-ins, emergencies, urgent patients) predetermined slots may be left open.

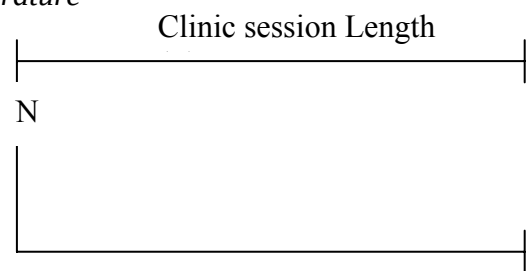
2.5 Appointment Rules

2.5.1 Appointment Rules Studied in the Literature

2.5.1.1. Single-Block

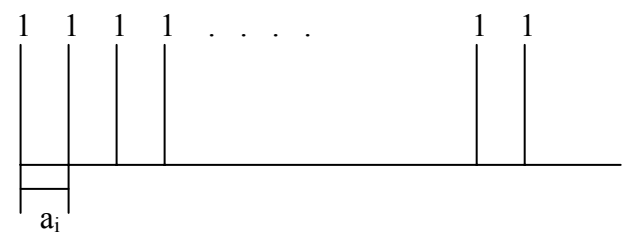
$$n_1 = N$$

$$\text{no } a_i$$



2.5.1.2. Individual-Block/Fixed-Interval

$$n_i = 1 \text{ for all } i = 1, 2, 3, \dots, N$$

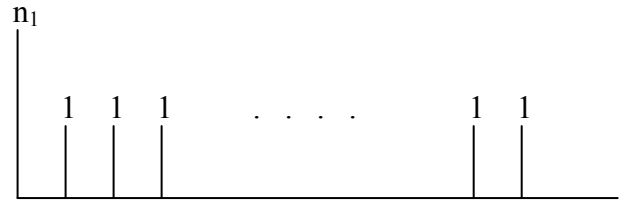


2.5.1.3. Individual-Block/Fixed-Interval with an initial block

$$n_1 > 1$$

$$n_i = 1 \text{ for all } i = 2, 3, \dots, N$$

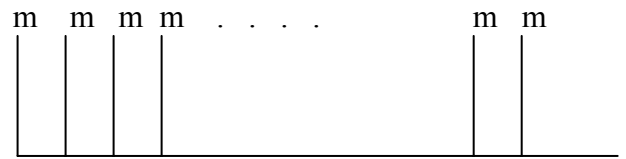
$$a_i \text{ constant}$$



2.5.1.4. Multiple-Block/Fixed-Interval

$$(m\text{-at-a-time})$$

$$n_i = m > 1 \text{ for all } i = 1, 2, 3, \dots, N$$

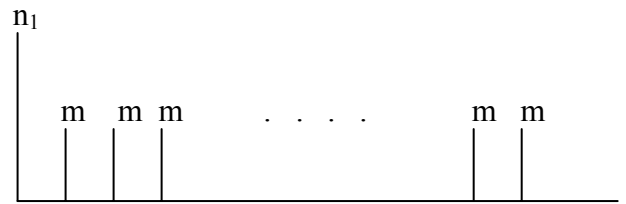


2.5.1.5. Multiple-Block/Fixed-Interval with an initial block

$$n_1 > m$$

$$n_i = m > 1 \text{ for all } i = 2, 3, \dots, N$$

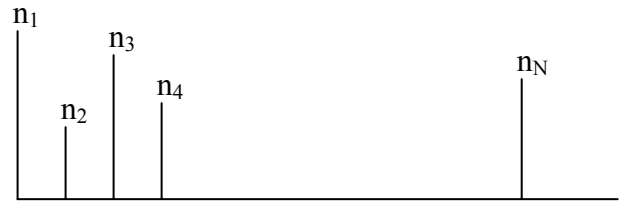
$$a_i \text{ constant}$$



2.5.1.6. Variable-Block/Fixed-Interval

$$n_i \text{ variable for all } i = 1, 2, 3, \dots, N$$

$$a_i \text{ constant}$$



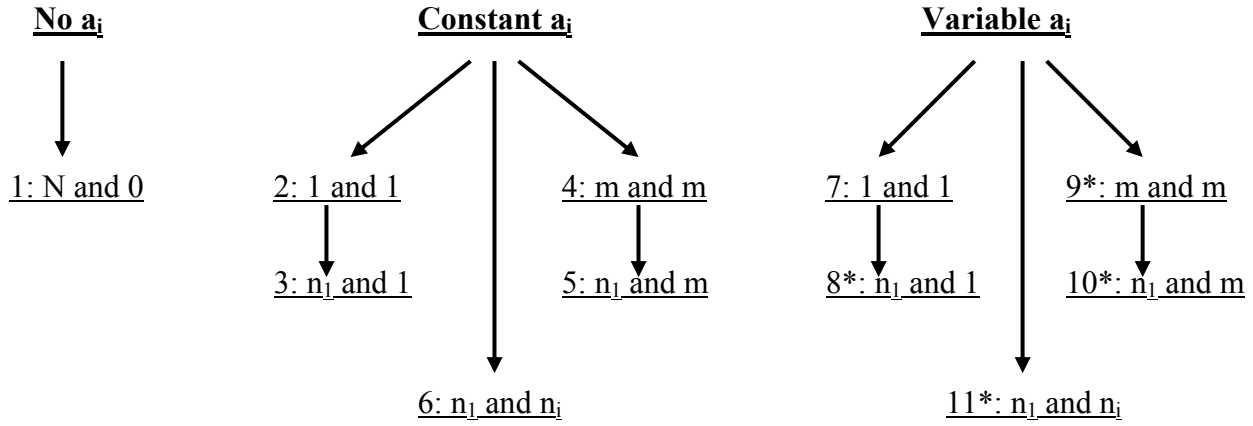
2.5.1.7 Individual-Block/Variable-Interval

$$n_i = 1 \text{ for all } i = 1, 2, 3, \dots, N$$

$$a_i \text{ variable}$$



2.5.2 Generalized Structure For Appointment Rules



Notification in underlined entries is as follows:

For: #: A and B:

#: the appointment rule number in the previous figure,

A: initial block

B: block size.

Rules 1-7 are studied in the literature while rules 8*-11* have not yet been studied in the literature.

2.6 Related Research Work

Cayırlı and Veral (2003) divide research methodologies in appointment scheduling literature into three classes: analytical, simulation-based or case study.

2.6.1 Analytical Approaches to Appointment Scheduling

The analytical approaches to the appointment scheduling studies include queuing theory and mathematical programming methods. Jansson (1966) studies a D/M/1 queuing model and obtains the total cost distribution function for the k^{th} customer. He assumes that patients are punctual and no-show probability is 0. By giving a C_d/C_p ratio, he determines the initial block and the constant appointment interval.

Soriano (1966) compares the waiting time distribution functions of individual and multiple-block/fixed interval appointment systems under the assumption of gamma service times and deterministic arrivals. He also assumes that all patients and doctors are punctual.

Fries and Marathe (1981) use dynamic programming approach to study variable-block/fixed interval appointments systems. They determine the optimal block sizes for the next period given that the remaining number of patients to be assigned is known. Then, they present a method to determine the schedule of a static version by using the dynamic programming results. They use theory of stochastic processes to determine the optimal block sizes and then by using dynamic programming they assign patients to the blocks. Liao et al. (1993) also use dynamic programming to determine the optimal block sizes given that service times are Erlang. Then they solve the static problem by a branch-and-bound algorithm and the solution of the dynamic programming is used as the lower bound of the algorithm. This algorithm is restricted to small-scale problems but Vanden Bosch et al. (1999) develop an efficient algorithm to solve the same but large-scale problems. Wang (1993) studies both dynamic and static states. He uses non-linear equations to find optimal appointment times. As a following study, Wang (1997) investigate the problem of scheduling a finite number of customer arrivals for a single-server system. A number of numerical examples are used to analyze the effect of the service time variations and the sensitivity of nondeterministic arrivals. A computational procedure is developed that efficiently evaluates the mean flow times and this procedure is integrated with a non linear program that determines the optimal customer arrival times. As a result, he concludes that the optimal appointment intervals are not constant. Liu and Liu (1998) develop a dynamic programming model to find the optimal block sizes and use the solution to determine the solution for static state. They compare different service time distributions such as uniform, exponential and weibull. They present two algorithms to determine block sizes under unknown patient population and by making some modifications they apply the dynamic programming model to static state problems.

Brahimi and Worthington (1991-a) study an appointment system with non-homogeneous arrivals and general discrete service time distributions. They present a Markov-chain algorithm to determine the distributions of number of patients in the system and analyze a number of scenarios to be presented to hospital management. Robinson and Chen (2001) present a stochastic linear program to find the optimal appointment times. The objective of the study is to minimize the expected weighted sum of the waiting time for the patients and the idle time of the doctor. By assuming that all patients and doctors are punctual, they develop a heuristic policy for setting job allowances and appointment times. They use generalized lambda distribution for service times which is fitted to data collected on surgery times. Denton and Gupta (2001) formulate a two-stage stochastic linear programming model to find optimal appointment times for a sequence of jobs with uncertain durations. Service times are illustrated for uniform, gamma and normal distributions. Mondschein and Weintraub (2003) present an economic framework to analyze the problem. Based on this framework they make a critical analysis of the objective functions used in the literature. They also question the validity of assumptions made throughout the literature. They conclude that the objective functions used in the literature are appropriate only in the case of a central planner that maximizes the social welfare. For other scenarios such as a private server, that maximizes his own utility, they are not appropriate. Then, they develop a more general model that fits these scenarios.

2.6.2 Simulation Approach to Appointment Scheduling

Cayırılı and Veral (2003) say that simulation has some advantages over analytical approaches such as ability to model complex systems and represent environmental variables.

Bailey (1952) is the first to investigate patients' and doctors' waiting time distributions in an outpatient medical facility. He suggests that in many medical clinics the ratio of waiting time to service time is too high. He argues against the single-block appointment system which is common in practice. This practice minimizes the doctor's idle time while maximizing the patients' waiting time. He

recommends an individual scheduling at regular intervals. He states that the longer the appointment interval, the higher the consultant's wasted time and the less the patients' waiting time. He fits Pearson type curves to the service distributions. He obtains waiting time distributions of patients by Monte-Carlo approach and searches for the best initial block. As a result he concludes that an initial block of two patients is reasonable for the balance between patient waiting time and doctor idle time. This is known as the "Bailey's rule" and it is widely studied in the literature. He assumes that all patients are punctual.

White and Pike (1964) use simulation to prevent excessive waiting times. They compare patient waiting time when patients are punctual or unpunctual and conclude that they do not differ greatly.

Katz (1969) develops a Monte Carlo simulation model of hospital outpatient scheduling to evaluate the effectiveness of alternative appointment systems. He takes into account lateness, multiple physicians and lab scheduling. In his model there are three kinds of inputs: simulation control inputs, clinic facilities and patient queues. The program is mechanized in FORTRAN as simulator. This simulator produces three kinds of reports: initial report, intermediate report and final report. As a result; the average waiting time of patients' and idle time of the doctor is obtained. Such a system can be applied for different appointment rules and they can be compared.

Vissers (1979) presents a heuristic to determine an appropriate appointment system to optimize patient waiting time and doctor idle time. He introduces a new variable "prepunctuality" and investigates the relationship between prepunctuality and patient waiting time. Ho and Lau (1992, 1995, and 1999) evaluate fifty different appointment rules and test their performance. They conclude that there is no specific appointment rule that can be effective under all circumstances and present a heuristic to select the best appointment rule. They also evaluate the impact of the environmental factors which include probability of no-show, the coefficient of variation and the number of patients per service session by using uniform and exponential service times with various coefficient of variation values. Klassen and Rohleder (1996) use simulation to compare alternative models based on variable

service times. Service times are distributed as a lognormal distribution and various coefficient of variation values are compared. They conclude that low-variance patients must be scheduled at the beginning of the session. As a later study, Rohleder and Klassen (2001) consider the possibility of schedulers' error to classify the patients because of increasing pressure to maximize utilization. In this later study, a service time variance is randomly chosen from a lognormal distribution.

Liu and Liu (1998) study a variable-block/fixed interval appointment system under consideration of doctors' lateness and three alternative service time distributions: uniform, exponential and weibull. They use simulation to determine optimal block size and present a simple procedure to select an appointment rule. Yang et al. (1998) build a mathematical formulation in terms of mean and standard deviation of service times that are distributed according to gamma distribution. They propose a new appointment rule for the single server system. They present the new formulation as a mathematical function and after determining the parameters the system is evaluated by simulation. To validate the effectiveness of new appointment rule, the new rule is tested and compared with eight best rules proposed by Ho and Lau.

Swisher et al. (2001) Use discrete event simulation to model a physician clinic environment within a physician network. The proposed physician network is composed of clinics located throughout the United States. The goal of the system is to increase patient flow and customer satisfaction. To provide a tool for decision making, a simulation model is developed. The input parameters of the model are composition of medical staff, number of registration windows, number of examination rooms, number of check-in rooms and number of specialty rooms. The collected statistics are about medical staff utilization, facility utilization, patient throughput, clinic overtime and patient time in the system. The model includes a standard clinic layout, standard medical personnel and standard patient types. The model is run for fifteen simulated months and as a result they conclude that the effect of number of check-in rooms is negligible and increasing the number of specialty rooms from 1 to 2 decreases the cost.

2.6.3 Case Studies in Appointment Scheduling

In case studies, researchers analyze a particular outpatient clinic, make recommendations for improving the existing system and evaluate the results of implementing these recommendations (Cayırlı&Veral, 2003).

Rising et al. (1973) use simulation to model a university outpatient clinic to improve patient and physician scheduling. They assume a negative exponential arrival time for walk-ins and emergencies. Their study results a decrease in clinic overtime for doctors and waiting time for patients. As a result they show that alternative decision rules for scheduling appointments can increase patient throughput and staff utilization. Mahacheck and Knabe (1984) present a case study in OB/GYN clinic of Johns Hopkins Hospital. They classify patients as new and return. They analyze alternative operational decisions such as patient scheduling methodology, personnel utilization, patient waiting time, staff arrival/departure. They use Poisson service times and the doctor lateness that is based on actual times observed. O'Keefe (1985) analyzes three outpatient departments in the U.K. He classifies patients as new and return. The aim of the study is to reduce patient waiting time. Cox et al. (1985) use simulation to evaluate an ear, nose and throat clinic. The objective of the study is to find the optimal appointment system to keep queue sizes constant and doctor work rates constant. They compare various ways of scheduling new / return patients by using various service time distributions.

Babes and Sarma (1991) investigate a single-block appointment system in some departments of the Ibn-Rochd health centre. After presenting input analysis and parameter estimation, they use discrete event simulation to determine the number of patients per clinic and the number of doctors to optimize the cost. They assume that service time distribution is Weibull. They also examined the sensitivity of performance parameters. Brahimi and Worthington (1991-b) apply their study to seven clinics in the U.K. They compare alternative systems based on a number of performance measures and as a result they conclude that an individual-block/fixed interval system with an initial block of three patients is optimal.

Huarng and Lee (1996) develop a simulation model to evaluate the effects of appointment system, staffing policies and service units. They build six models with different sessions, different number of physicians and different number of cashiers. They use SLAM as a simulator and as a result they conclude that a great decrease in patient waiting time and workload of the doctor is possible by changing arrival process or service process. If more patients are scheduled by appointment, time in the system decreases. Also the results show that finding another session for the physician practicing in dermatology will increase the number of patients treated. Lehaney et al. (1999) use simulation to reduce unexpected non-attendance of patients. The main suggestion they make is that patient appointment times should be front-loaded to reduce queues and reduce the possibility of idle time for doctors because of non-attending patients.

Vanden Bosch and Dietz (2000) analyze scheduling and sequencing policies for a specific clinic with the goal of reducing combination of patient waiting time and doctor idle time. They classify patients according to their appointment history or procedure type. They use an analytical approach to solve static state problem under the assumptions of punctuality and stochastic service times. As a result, substantial reductions in waiting times can be obtained. After establishing the optimization policy, they present Lattice algorithm that is easy to implement. The objective of the study of Su and Shih (2003) is analyzing several scheduling solutions by simulation and finding the appropriate arrival time interval. He built four alternative models of different arrival intervals. Patients may be registered or walk-ins with different probabilities of existence. The first model is the current state of the system in which registered patients are prior. The second model proposes a scheduling system of scheduling front numbers for registered patients and later numbers for walk-ins. In the third model, odd numbers are reserved for registered patients while odd numbers are for walk-ins. The last model is solved by two stages. First, among different time intervals the optimum one is found and then by using this optimum interval performance measures are calculated.

Harper and Gamlin (2003) analyzed an ear, nose and throat outpatient department by using simulation. Alternative appointment schedules are evaluated to reduce patient waiting time without using extra resources and under the restriction of The Patient's Charter: all patients must be seen within thirty minutes.

Consequently, there are various approaches to solve appointment scheduling problems. An overview of commonly used techniques is given in the next section. Since dynamic programming is used as the solution methodology in this thesis, it is described in detail.

CHAPTER THREE

OVERVIEW OF SOLUTION METHODOLOGIES USED FOR APPOINTMENT SCHEDULING

Research methodologies used for appointment scheduling are:

- 1) Simulation
- 2) Stochastic linear programming
- 3) Dynamic programming

3.1 Simulation

A simulation is the imitation of the operation of a real-world process or system over time. Whether done by hand or on a computer, simulation involves the generation of an artificial history of a system. The behavior of a system is studied by developing a simulation model. This model usually takes the form of a set of assumptions concerning the operation of the system. Once developed and validated, a model can be used to investigate a wide variety of “What-If” questions about the real-world system. Potential changes to the system can be simulated in order to predict their impact on system performance. Simulation can also be used to study systems in the design stage, before building the system. Thus, simulation modeling can be used both as an analysis tool for predicting the effect of changes to existing systems and as a design tool to predict the performance of new systems under varying sets of circumstances.

In some cases, a model can be developed which is simple enough to be solved by mathematical methods. Such solutions may be found by the use of differential calculus, probability theory or other mathematical techniques. The solution usually consists of one or more numerical parameters as performance measures of the system. However, many real-world systems are so complex that models of these systems are virtually impossible to solve mathematically. In these cases, numerical, computer-based simulation can be used to imitate the behavior of the system.

3.1.1 When Simulation Is Appropriate?

The availability of special purpose simulation languages, massive computing capabilities at a decreasing cost per operation and advances in simulation methodologies have made simulation one of the most widely used and accepted tools in operations research and systems analysis. Simulation can be used for the following purposes:

- 1) Simulation enables the study of the internal interactions of a complex system, or of a subsystem within a complex system.
- 2) Informational, organizational and environmental changes can be simulated and the effect of these alterations on the model's behavior can be observed.
- 3) By changing simulation inputs and observing the resulting outputs, valuable insight may be obtained into which variables are most important.
- 4) Simulation can be used to experiment with new designs or policies prior to implementation.
- 5) Simulation can be used to verify analytic solutions.
- 6) Animation shows a system in simulated operation so that the plan can be visualized.
- 7) The modern system (factory, service organization etc.) is so complex that the interactions can be treated only through simulation.

3.1.2 When Simulation Is Not Appropriate?

Based on the article by Banks and Gibson (1997), simulation is not an appropriate tool for such cases:

- 1) Simulation should not be used when the problem can be solved using common sense.
- 2) Simulation should not be used if the problem can be solved analytically.
- 3) Simulation should not be used if it is easier to perform direct experiments.

- 4) Simulation is not appropriate if the costs exceed the savings.
- 5) If there is not enough time or the personnel are not available to verify and validate the model, simulation is not appropriate.
- 6) If the system behavior is too complex or cannot be defined, simulation is not appropriate.

3.1.3 Advantages And Disadvantages Of Simulation

In contrast to optimization models, simulation models are “run” rather than solved. Given a particular set of input and model characteristics, the model is run and the simulated behavior is observed. This process of changing inputs and model characteristics results in a set of scenarios that are evaluated. A good solution, either in the analysis of an existing system or the design of a new system, is then recommended for implementation.

According to Pegden, Shannon and Shadowski (1995) simulation has many advantages and even some disadvantages.

The advantages of the simulation are as follows:

- 1) New policies, operating procedures, decision rules, information flows, organizational procedures and so on can be analyzed without interrupting ongoing operations of the real system.
- 2) New hardware designs, physical layouts, transportation systems, and so on can be tested without committing resources.
- 3) Hypothesis about how or why certain phenomena occur can be tested for feasibility.
- 4) Time can be compressed or expanded allowing for a speedup or slowdown.
- 5) Insight can be obtained about the importance of variables to the performance of the system.

- 6) Bottleneck analysis can be performed indicating where processes are being excessively delayed.
- 7) “What-If” questions can be answered. This is particularly useful in the design of new systems.
- 8) Simulation is applicable in complex cases where analytical procedures cannot be employed.

The disadvantages of the simulation are:

- 1) Model building requires special training. It is an art that is learned over time and through experience. This disadvantage has been reduced with the development of computer simulation languages such as GPSS, SIMSCROPT, SIMAN and SLAM.
- 2) Simulation results may be difficult to interpret.
- 3) Simulation modeling and analysis can be time consuming and expensive.
- 4) Simulation does not guarantee an optimal solution to a problem (Banks et al., 2001).

3.2 Stochastic Linear Programming

Stochastic programming is a framework for modeling optimization problems that involve uncertainty. Whereas deterministic optimization problems are formulated with known parameters, real world problems almost include some unknown parameters. When the parameters are known robust optimization is used to find a solution which is feasible for all such data and optimal in some sense. Stochastic programming models are similar in style but take advantage of the fact that *probability distributions* about the data are known or can be estimated. The goal of these models is to find some policy that is feasible for all (or almost all) the possible data instances and maximizes the expectation of some function of the decisions and the random variables. More generally, such models are formulated, solved analytically or numerically, and analyzed in order to provide useful information to make a decision.

The most widely applied and studied stochastic programming models are two-stage linear programs. In these programs decision maker takes some action in the first stage, after which a random event occurs, affecting the outcome of the first-stage decision. A *recourse* decision can then be made in the second stage that minimizes any bad effects that might have been experienced as a result of the first-stage decision. The optimal policy from such a model is a single first-stage policy and a collection of recourse decisions (a decision rule) defining which second-stage action should be taken in response to each random outcome.

Applications of stochastic linear programming are widespread, such as finance, capacity planning, agriculture, forestry, military, production control, scheduling and transportation.

3.2.1 Formulation of a Stochastic Linear Program

Stochastic programs seek to minimize the cost of the first-period decision plus the expected cost of the second-period recourse decision.

$$\begin{aligned} \min \quad & c^T x + E_w Q(x, w) \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

where

$$\begin{aligned} Q(x, w) = \min \quad & d(w)^T y \\ \text{subject to} \quad & T(w)x + W(w)y(w) = h(w) \end{aligned}$$

The first linear program minimizes the first-period direct costs, $c^T x$ plus the expected recourse cost, while meeting the first-period constraints, $Ax = b$

The recourse cost Q depends both on x , the first-period decision, and on the random event, w . The second linear program describes how to choose $y(w)$ (a different decision for each random scenario w). It minimizes the cost $d^T y$ subject to

some recourse function, $T\mathbf{x} + \mathbf{W}\mathbf{y} = \mathbf{h}$. This constraint can be thought of as requiring some action to correct the system after the random event occurs. One important thing to notice in stochastic programs is that the first-period decision, $\mathbf{x}$, is independent of which second-period scenario actually occurs.

3.3 Dynamic Programming

In this thesis, dynamic programming is used as the solution methodology of appointment scheduling problem in DEU, Dept.of Ort. and Trau. Since, the professors, associate professors and assistant professors treat patients on an appointment based system, these appointments are scheduled according to variable-sized, multiple-block appointment rule by using dynamic programming. Then, more detailed overview of dynamic programming is given here.

Dynamic programming is a useful mathematical technique for making a sequence of interrelated decisions. It provides a systematic procedure for determining the optimal combination of decisions (Hillier & Lieberman, 1995). This technique evolved as a result of studying programming problems in which changes over time are important and this is why it was given the name “dynamic programming”. However, the technique can be applied to problems in which time is in no way relevant.

Dynamic programming is an approach to problem solving that permits decomposing one large mathematical model that may be very difficult to solve into a number of smaller problems that are usually much easier to solve. Moreover, the dynamic programming approach allows breaking up a large problem in such a way that once all the smaller problems have been solved, an optimal solution to the large problem is left only. Each of the smaller problems created is identified with a stage of the dynamic programming solution procedure. As a consequence, the technique has been applied to many decision problems that are multistage in nature. Often the multiple stages are created by the fact that a sequence of decisions must be made over time. For example, a problem of determining an optimal decision over a 1-year

time horizon might be broken into 12 smaller stages, where each stage requires an optimal decision over a 1-month time horizon. In most cases each of these smaller problems cannot be considered to be completely independent of the others, and this is where the dynamic programming approach is helpful.

In contrast to linear programming, there does not exist a standard mathematical formulation of the dynamic programming problem. Rather, dynamic programming is a general type of approach to problem solving and the particular equations used must be developed to fit each situation.

To form a dynamic programming model, states of incomplete solution must be characterized to capture the history on which each new decision should be based. States in dynamic programming characterize conditions of incomplete solution at which decisions should be considered. If an optimal solution is known for any state, those of subsequent states can build upon it.

A sequence of decisions based on states of incomplete solution lies at the heart of all discrete dynamic programs. However, it is often convenient to distinguish stages of decision making from states of solution. The *stages* of dynamic programming solution procedure are formed by decomposing the original problem into a number of sub problems. In dynamic programs with both stages and states, stages indicate the sequence of required decisions and states encode the conditions within which decisions can be considered.

In dynamic programming, the decision maker's goal is usually to minimize expected (or expected discounted) cost incurred or to maximize expected (or expected discounted) reward earned over a given time horizon.

$$f_t(\text{current state}) = \min_{\text{all feasible decisions}}(\text{or max})\{\text{costs during current stage} + f_{t+1}(\text{new stage})\}$$

3.3.1 Characteristics Of Dynamic Programming Problems

The basic features that characterize dynamic programming problems are as follows:

- 1) The problem can be divided into “*stages*”, with a policy decision required at each stage. Dynamic programming problems require making a sequence of interrelated decisions, where each decision corresponds to one stage of the problem.

In appointment scheduling problems; “*stage*” refers to “*period*”. A clinic session is divided into K periods of equal length.

- 2) Each stage has a number of “*states*”, associated with the beginning of that stage. The states are the various possible conditions in which the system might be at that stage of the problem. The number of states may be either finite or infinite.

In appointment scheduling problems; the number of patients in the system at the beginning of each period is referred as “*state*” and the number of states is finite.

- 3) The effect of policy decision at each stage is to transform the current state to a state associated with the beginning of the next stage.

The number of patients assigned to a period affects the decision of next period during scheduling appointments in the dynamic state.

- 4) The solution procedure is designed to find an optimal policy for the overall problem. Dynamic programming provides a kind of policy prescription of what to do under every possible circumstance (which is why the actual decision made upon reaching a particular state at a given stage is referred to as a “*policy*” decision.)

A table is constructed for each stage that prescribed the optimal decision for each possible state.

- 5) Given the current state, an optimal policy for the remaining stages is independent of the policy decisions adopted in previous stage. The optimal immediate decision depends on only the current state and not on how to get there. This is the “**Principle of Optimality**” for dynamic programming.
- 6) The solution procedure begins by finding the optimal policy for the last stage. The optimal policy for the last stage indicates the optimal policy decision for each of the possible states at that stage.
- 7) A “**recursive relationship**” that identifies the optimal policy for stage n , given the optimal policy for stage $n+1$, is available. Recursive relationship:

$$f_n^*(s_n) = \min_{x_n} f_n(s_n, x_n) = f_n(s_n, x_n^*)$$

$$f_n^*(s) = \min_{x_n} (c_{sx_n} + f_{n+1}^*(x_n))$$

c_{sx_n} = immediate cost

$f_{n+1}^*(x_n)$ = min future cost

N : number of stages

n : label for current stage ($n: 1, 2, \dots, N$)

s_n : current state for stage n

x_n : decision variable for stage n (immediate decision)

x_n^* : optimal value of x_n (given s_n)

$f_n(s_n, x_n)$: contribution of stages $n, n+1, \dots, N$ to objective function if system starts in state s_n at stage n , immediate decision is x_n and optimal decisions are made thereafter.

$f_n^*(s_n)$ is defined in terms of $f_{n+1}^*(s_{n+1})$ that makes the expression for

$f_n^*(s_n)$ a recursive relationship.

The recursive relationship keeps recurring as moved backward stage by stage. When the current stage number n is decreased by 1, the new $f_n^*(s_n)$ function is derived by using the $f_{n+1}^*(s_{n+1})$ function that was just derived during the preceding iteration and then this process keeps repeating.

- 8) When the recursive relationship is used, the solution procedure starts at the end and moves backward stage by stage –each time finding the optimal policy for that stage- until it finds the optimal policy starting at the initial state (Hillier & Lieberman, 1995).

3.3.2 Deterministic Dynamic Programming

In deterministic dynamic programming problems, the state at the next stage is completely determined by the state and policy decision at the current stage. Deterministic dynamic programming can be described diagrammatically as follows:

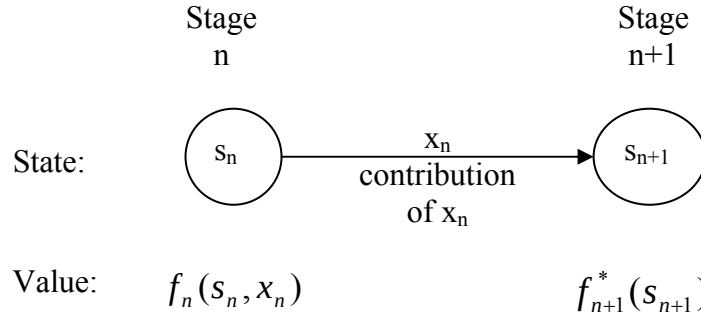


Figure3.1 The basic structure for deterministic dynamic programming

At stage n the process will be in some state s_n . Making policy decision, x_n then moves the process to some state s_{n+1} at stage $n+1$. The contribution to the objective function under an optimal policy has been previously calculated to be $f_{n+1}^*(s_{n+1})$. The policy decision x_n also makes some contribution to the objective function.

One way of categorizing deterministic dynamic programming problems is by the form of the objective function. For example; the objective might be to minimize the sum of the contributions from the individual stages or to maximize such a sum, or to minimize a product of such terms and so on. Another categorization is in terms of the nature of the set of states for the respective stages. In particular, states s_n might be representable by a discrete state variable or by a continuous state variable or perhaps a state vector is required.

3.3.3 Probabilistic Dynamic Programming

Probabilistic dynamic programming differs from deterministic dynamic programming that the state at the next stage is not completely determined by the state and policy decision at the current stage. In probabilistic dynamic programming there is a **“probability distribution”** for what the next state will be. This probability distribution is completely determined by the state and policy decision at the current stage. Probabilistic dynamic programming can be described diagrammatically as follows:

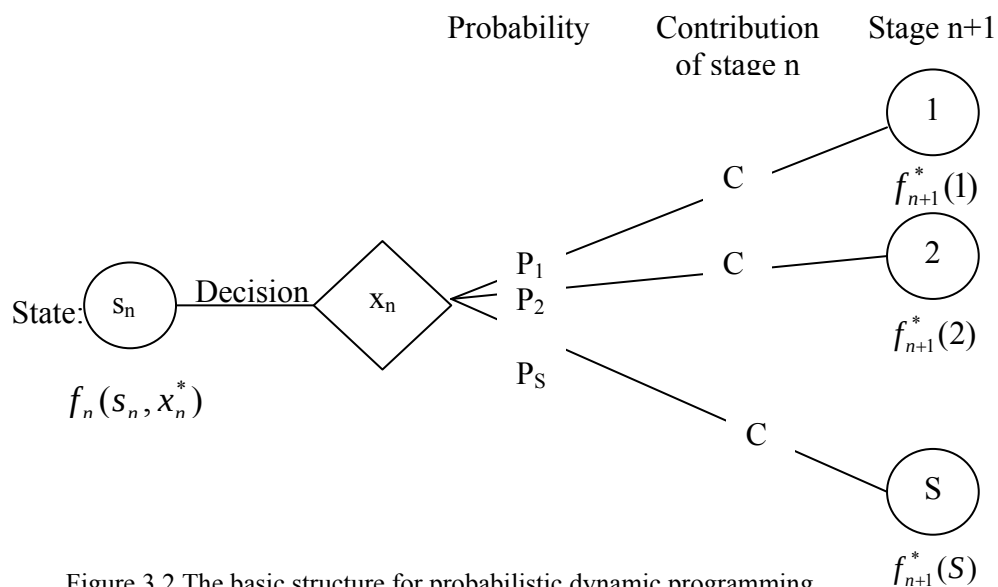


Figure 3.2 The basic structure for probabilistic dynamic programming

“S” denotes the number of possible states at stage $n+1$. The system goes to state i with probability p_i . If the system goes to state i , c_i is the contribution of stage n to the objective function.

Because of the probabilistic structure, the relationship between $f_n(s_n, x_n)$ and $f_{n+1}^*(s_{n+1})$ is more complicated than that for deterministic dynamic programming.

$$f_n(s_n, x_n) = \sum_{i=1}^S p_i [C_i + f_{n+1}^*(i)]$$

with

$$f_{n+1}^*(i) = \min_{x_{n+1}} f_{n+1}(i, x_{n+1})$$

Since the system analyzed in this thesis has a probabilistic structure, the above formulations will be used to solve the problem.

3.3.4 Dynamic Programming In Appointment Scheduling

The key paper that will be modified and applied to solve dynamic programming model of scheduling appointments of a doctor in DEU, Dept.of Ort. and Trau. is as follows:

“Planning Timely Arrivals to a Stochastic Production or Service System”, Liao & Pegden & Rosenshine(1993), Journal of Operational Research Society, 50, 877-891

This study is concerned with N customers require service at a stochastic system. The service times have either an exponential or Erlang distribution. A finite time interval is divided into “ k ” equal time slots and each customer is to be scheduled to arrive at the beginning of one of the time slots. The objective of the study is to determine how many of these N customers should be scheduled to arrive at each time slot to minimize the sum of the customer waiting cost and server availability cost. In other words the problem is to balance the total customer waiting cost against the server overtime cost.

As the first step, dynamic version solution is presented and then a branch-and-bound algorithm is developed for solving the static state.

Dynamic Version:

A dynamic programming solution for the problem is presented with such notations:

r : remaining customers unscheduled at the beginning of stage i ,

d_i : customers assigned to stage i ,

s : the number of previously scheduled customers in the system

Then, there will be $r - d_i$ customers remaining to be assigned at the beginning of stage $i+1$.

Then the optimal value function $f_i(r, s)$ is defined as “the minimum expected cost obtainable due to unassigned customers from time slots i through k given r customers remaining to be assigned and s previously scheduled customers still in the system”.

$$f_i(r, s) = \min_{0 \leq d_i \leq r} g_i(r, s, d_i) \quad (1)$$

$$= \min_{0 \leq d_i \leq r} \left\{ \frac{c_w d_i}{2\mu} (2s + d_i - 1) + \left[\sum_{j=0}^{s+d_i} P(s + d_i - j | s + d_i) * f_{i+1}(r - d_i, s + d_i - j) \right] \right\} \quad (2)$$

$i = 1, 2, \dots, k$

μ : mean service time

c_w : customer waiting cost per unit time

$P(m - j | m)$: Transition probability of moving from state m to state $m-j$.

$$P(m - j | m) = p(j) \quad j \leq m - 1 \quad (3)$$

$$P(0 | m) = 1 - \sum_{j=0}^{m-1} p(j) = p(\bar{m}) \quad (4)$$

$p(j)$: probability of j service completions in one time slot.

The boundary condition of the model is that in the last period all remaining customers must be scheduled:

$$\begin{aligned} f_k(r, s) &= g_k(r, s, r) \\ &= \frac{c_w r}{2\mu} (2s + r - 1) + c_s \sum_{j=0}^{r+s-1} P(r + s - j | r + s) \frac{r + s - j}{\mu} \end{aligned} \quad (5)$$

c_s : server overtime cost per unit time

Since the server leaves if there is no remaining customer there is no server availability cost for the last period.

In (3) m is the number of customers in the system after d_i customers are assigned into time slot i while $m-j$ is the number of customers before the assignment d_{i+1} . Hence; $P(m-j | m)$ is the probability that exactly j customers are actually served during the time period of one slot. The recursive relation (2) is established by noting that there are r remaining customers at the beginning of stage i so any number of customers from 0 to r can be assigned to this stage. If d_i customers were assigned, the j^{th} customer of d_i would need to spend, on the average, $(s+j-1)/\mu$ units of time in the queue, due to the Markovian property of the exponential service times. Thus, the total waiting cost for these d_i customers will be:

$$c_w \sum_{j=1}^{d_i} \frac{s+j-1}{\mu} = \frac{c_w}{\mu} \left[d_i(s-1) + \frac{d_i(d_i+1)}{2} \right] = \frac{c_w d_i}{2\mu} (2s + d_i - 1) \quad (6)$$

The remaining terms in (2) are the expected cost for the subsequent stages $i+1$ through k . It is a weighted average of f_{i+1} 's with $r-d_i$ customers remaining to be scheduled in the subsequent stages and $s+d_i-j$ customers in the system at the beginning of stage $i+1$. The weights are the transition probabilities associated with the completion of j services during stage i ($j=0,1,2,\dots,s+d_i$).

It is noted that the problem solution depends only on the ratio of c_s to c_w and not on their individual values. In practice, the individual values may be difficult to obtain while their ratio may be easier to estimate.

If $P_i(r, s)$ is the optimum decision for state (r, s) at stage i , then the following theorems about the dynamic optimal policy may be proved:

Theorem 1:

For all $i, r \geq 1$, and $s \leq N - 1$,

$$\text{if } P_i(r, s) \geq 1, \text{ then } f_i(r - 1, s + 1) = f_i(r, s) - c_w \frac{s}{\mu}$$

Theorem 2:

$$P_i(r, 0) \geq 1 \text{ for all } i \text{ and } r \geq 1$$

Theorem 3:

$$P_{i+1}(r, s) \geq P_i(r, s) \text{ for } i \leq k - 1, \text{ all } r \text{ and } s$$

Theorem 4:

$$P_i(r, s) - P_i(r - 1, s) = 0 \text{ or } 1 \text{ for all } i, s \text{ and } r \geq 1$$

Theorem 5:

$$P_i(r, s) = (P_i(r + 1, s - 1) - 1)^+ \text{ for all } i, r \leq N - 1 \text{ and } s \geq 1$$

Based on the results, without using (2) and (5) to obtain $f_i(r, s)$, the following short-cut method is developed:

$$f_i(1, 0) = g_i(1, 0, 1) \quad i \leq k - 1 \quad (7)$$

$$f_i(r, 0) = \min_{P_i(r-1,0) \leq d_i \leq \min\{P_i(r-1,0)+1, P_{i+1}(r,0)\}} g_i(r, 0, d_i) \quad i \leq k - 1, r \geq 1 \quad (8)$$

$$f_i(r, s) = f_i(r + 1, s - 1) - c_w \frac{s - 1}{\mu} \quad P_i(r + 1, s - 1) \geq 1, \text{ all } i, r \leq N - 1, s \geq 1 \quad (9)$$

$$P_i(r + 1, s - 1) = 1, \quad i \leq K - 1, \text{ all } r, s \geq 1 \quad (10)$$

where function g 's are the same as in (2) and (5). For $i=k, s=0$ still (5) is used.

Static Version:

It is known that each decision in the dynamic version is based on what has happened during the previous time period and need not even be specified until the previous time period is completed. However, in the static version all decisions are made prior to the start of the first time period.

First, the lower bounds that are calculated by using following formulas:

$$LB(\underline{d}_i) = \sum_{s=0}^{D_{i-1}} v_i(s) g_i(N - D_{i-1}, s, d_i) + h_{wc}(\underline{d}_{i-1}) \quad (11)$$

Where

$$v_i(s) = \sum_{j=(s-d_{i-1})^+}^{D_{i-2}} v_{i-1}(j) P(s | d_{i-1} + j) \quad s = 0, 1, \dots, D_{i-1} \quad (12)$$

$$h_{wc}(\underline{d}_{i-1}) = c_w \sum_{s=0}^{D_{i-1}} v_{i-1}(s) \sum_{j=1}^{d_{i-1}} \frac{s + j - 1}{\mu} + h_{wc}(\underline{d}_{i-2}) \quad (13)$$

The following notation used in this formulation is:

$\underline{d}_i$ = sequence of decisions from stages 1 through i. ($\underline{d}_i = (d_1, d_2, \dots, d_i)$);

$$D_{i-1} = \sum_{j=1}^{i-1} d_j ;$$

$h_{wc}(\underline{d}_i)$ = total expected waiting cost from customers scheduled in stages 1 through i;

$LB(\underline{d}_i)$ = lower bound on the expected costs of those schedules with $\underline{d}_i$ as the decisions from stages 1 through i;

S = number of previously scheduled customers remaining in the system at the beginning of the stage (a random variable);

$v_i(s)$ = probability density function of the random variable S at stage i;

$(j)^+ = \max \{0, j\}$;

where any D , $\underline{d}$, v or h_{wc} with a non-positive subscript is defined to be zero.

A new vector $\bar{d}_i$ denotes the decisions from stages i through k . $\bar{d}_i = (d_i, d_{i+1}, \dots, d_k)$ for $i=1, 2, \dots, k$. It is necessary to determine the minimum expected cost from stages i through k for any schedules with d_i being made. Since $\bar{d}_i$ contains the decisions d_i through d_k that have not been decided at stage i , calculations of dynamic version is used to determine this cost. Also the S at stage i a random variable for given $\underline{d}_i$ so a weighted average of g_i is used and the weights are the probabilities $v_i(s) = P(S=s)$. This explains the first term in (11). The second term is the total expected waiting cost from customers scheduled in stages 1 through $I-1$ and is given by (13).

Then, branch-and-bound approach is used to find the static solution.

CHAPTER FOUR

OVERVIEW OF PERSONNEL SCHEDULING

4.1 Personnel Scheduling

In this thesis, personnel scheduling is used to schedule the working hours of residents in DEU, Dept.of Ort. and Trau. In the department there are 24 residents that work in the outpatient clinic. Linear programming is used for scheduling work hours of these residents and an overview of linear programming approach is given in the next section.

Personnel scheduling deals with the arrangement of work schedules and the assignment of personnel to shifts to cover the demand for resources that vary over time. These problems arise in manufacturing as well as in services. Effective personnel scheduling is one of the critical success factors for companies operating in dynamic environments such as restaurants and catering organizations, hospitals (scheduling shifts for doctors and nurses), educational institutions (for temporary teaching needs), trucking and railroad companies, and so on. Without sufficient numbers of doctors and nurses, hospitals cannot properly manage patient care; educational institutions without teachers cannot provide effective instruction. In service environments, the operations are often prolonged and irregular and the staff requirements fluctuate over time. The schedules are typically subject to various constraints dictated by equipment requirements, union rules and so on (Pinedo &Chao, 1999).

In many industries, the demand for services varies considerably over time. For example; a mail-order company needs customer service telephone operators on a 24-hour basis, 365 days per year. However; the number of operators needed may vary by the hour, day of the week and even the month of the year. It would be too expensive to maintain a full-time work force at a staffing level to meet peak period demands during the holiday ordering season. As a result, many companies hire part-time employees (Avriel&Golany, 1996). On the other hand, for periods with less

demand level, it is also expensive to have all full-time work force scheduled. As a result, it is necessary to make schedules to meet demand of each period. These problems are called shift scheduling or staff planning problems. In shift scheduling or staff planning models, the work is already fixed and the resources to accomplish it must be planned. In particular, how many of what types of workers and shifts best cover all work requirements are determined.

At the same time, organizations with varying personnel requirements need to take into consideration not only the workers' availability, but must schedule workers based on other criteria such as their skills set. Traditionally, scheduling activities have been performed manually, causing companies to struggle with expensive and time-consuming processes. Automation of personnel scheduling does not only provide a cost effective solution, but it also presents a more reliable scheduling method. This, in turn, becomes a key competitive differentiator for service organizations in a growing market.

The most common variations of personnel scheduling problems are days-off scheduling and shift scheduling. Days-off scheduling is an elementary personnel assignment. Each day of week, a number of employees have to be present and these numbers may differ from day to day while the set of requirements remains the same from week to week. There is a total number of employees available and each has to be assigned a sequence of days. In shift scheduling problems, a cycle that is fixed in advance is considered. In certain settings the cycle may be a single day, whereas in others it may be a week or a number of weeks. The decision variables of the model give the number of people assigned to a specific shift. The objective of the model is to minimize the total cost of assigning personnel to meet demand.

4.2 Literature Review

Personnel scheduling algorithms can simply be classified according to the techniques used to solve as heuristics, self-scheduling and mathematical programming (Gorantla, 1995).

4.2.1 Heuristic Approaches to Personnel Scheduling

Megeath (1978), proposes a scheduling algorithm concerned with days-off problem. After determining days-off specific shifts are assigned to comply with organizational work patterns. Maier-Rothe and Wolfe (1973), propose a heuristic based algorithm to distribute unpopular shifts by selecting days-off patterns. They build schedules to ensure approximate mix of specialties and the minimum coverage.

Smith and Wiggins (1977), incorporates work patterns, vacations, holidays and special requests in their algorithm. They build alternative days-off sequences that meet staff requirements and schedules are developed for each individual. If a particular individual cannot be scheduled, a work pattern constraint is relaxed and alternative schedules are generated. Smith (1976), analyzes the influence of work patterns on scheduling process. First, he determines how many staff members are required for the necessary coverage and then proposes an algorithm for the staff size that accommodates both days-off policies and coverage requirements. This algorithm also allows comparison of relative costs. Blau (1985) uses request points to distribute unpopular work shifts. Individual dissatisfaction scores are calculated to measure how often a staff member has not received his/her special request.

In general, heuristic personnel scheduling algorithms require identification of the acceptable schedules that meet the constraints of the organization. They can be identified by various means including determining days-off and then scheduling the individual shifts that provide a staff size equal to or greater than the predicted demand (Gorantla, 1995).

4.2.2 Self-Scheduling

Miller (1984), Ringl and Dotson (1989) describe personnel scheduling approaches referred as self-scheduling. Using this approach, the supervisor or the scheduler works with employees to determine how many employees are needed for

each shift. Staff members select which shifts they want to work in a manner consistent with previously agreed-upon organizational policies and procedures. The scheduler reviews the proposed self-schedules and modifies them as appropriate to meet the organizational requirements. Self-scheduling requires close cooperation and support by the staff (Gorantla, 1995).

4.2.3 Mathematical programming Approaches to Personnel Scheduling

Mathematical programming techniques used to schedule personnel are generally based on optimization concepts of linear programming. Numerous objective functions are used in the literature such as minimizing the cost (Bailey and Fields, 1985), minimizing the number of employees to be scheduled (Rosenbloom and Goertzen, 1987), minimizing staff dissatisfaction (Arthur and Ravindran, 1981), and minimizing penalty for exceeding staff levels (Miller, Pierskalla and Rath, 1976). Each of the objective functions is subject to constraints in order to provide the necessary coverage. When constraints cannot be made the scheduler is required to relax constraints by accepting short staffing or increasing the staff size.

Wolfe and Wolfe and Young (1965) are the first to use linear programming in personnel scheduling. The objective function was minimizing a combination of salary cost and a utility factor for performing all tasks in a shift but they do not constrain the problem by limiting the staff size.

Ozkarahan and Bailey (1988) propose a goal programming approach to the personnel scheduling. The goals of the model are providing the necessary coverage and scheduling the staff for their contracted work hours. They suggest that the problem can be divided into two stages. First, goal programming is performed to determine alternative schedules for day-of-week and time-of-day problems. Then, individuals are assigned to the proposed scheduled heuristically.

Bartholdi, Orlin and Ratliff (1980) use integer programming to solve the problem. According to the problem, employees work five out of seven days individual preferences are not considered. Bailey and Field (1985) use linear programming to identify feasible shift configurations to minimize cumulative costs. The identified schedules are either offered to the staff for self-scheduling or assigned using a linear program but they are not interested in individual preferences.

To summarize, mathematical programs use linear, integer or goal programming techniques to identify feasible sets of schedules but individual assignment is generally done by using heuristic approaches.

Since, linear programming approach is used as the solution methodology of personnel scheduling problem in this thesis, an overview of the technique is given in the next section.

4.3 Solution Methodology for Personnel Scheduling

4.3.1 Linear Programming

In this thesis, linear programming is used for solving personnel scheduling problem in DEU, Dept.of Ort. and Trau. There are 24 research assistants in the department that treat patients in the outpatient clinic. The working hours of these doctors are scheduled by using linear programming approach.

A linear program is an optimization model in which the objective and the constraints are linear in the variables to be determined. That is, the model's single-objective function and all constraints must consist of weighted sums of continuous decision variables. A linear program can be expressed as follows:

$$\begin{array}{ll}
\min & \bar{c}\bar{x} \\
\text{subject to} & \\
& A\bar{x} \leq \bar{b} \\
& \bar{x} \geq 0
\end{array}$$

where

$\bar{c}$: Cost vector

$\bar{x}$: Vector of decision variables

A : Activity vector

$\bar{b}$: Resource vector

There are several algorithms or classes of algorithms for solving linear programs. The most important ones are simplex method and interior point methods. Simplex method works very well in practice but there is not a version of the simplex method that solves linear programs in polynomial time. The most known example of an interior point method is the Karmarkar algorithm, which is known to solve the linear program in polynomial time.

The linear program built in the study for personnel scheduling is based on the following article. The formulations are modified and applied to the personnel scheduling problem in DEU, Dept.of Ort. and Trau. outpatient clinic. To solve the linear programs, a software, LINGO is used. An overview of the software is given in Appendix H.

4.3.2 “An Integrated Nurse Scheduling Model”, Ozkarahan, Irem, 1991

This study is concerned with scheduling shifts of personnel to satisfy the interests of both hospitals and nurses. An integrated scheduling model is built that accommodates flexible work patterns by incorporating time-of-day and day-of-week scheduling problems. Bailey’s (1985) integration formulation for limited staff size is the following:

$$\min Z = \sum_{j=1}^J C_j X_j + \sum_{i=1}^7 \sum_{l=1}^{L_i} \alpha_{il} U_{il}$$

subject to

$$\sum_{j=1}^J A_{ij} X_j - \sum_{k=1}^K X_{ik} = 0 \quad \text{for } i = 1, 2, \dots, 7$$

$$\sum_{k=1}^K A_{ik} X_{ik} - O_{il} + U_{il} = D_{il} \quad \text{for } l = 1, \dots, L \text{ and } i = 1, \dots, 7$$

$$\sum_{j=1}^J X_j = W_{total}$$

$$X_{ik}, X_j \geq 0 \quad \text{and integer}$$

where

A_{ik} = an L_i by K_i matrix where $a_{ik}=1$ if pattern k is working during period l and $a_{ik}=0$ otherwise

D_{il} = demand for workers during period l on day i ;

K_i = number of feasible pattern on day i ;

L_i = number of periods to be scheduled on day i ;

X_{ik} = number of workers scheduled to work shift pattern k on day i ;

O_{il} = overstaffing surplus during period l of day i ;

U_{il} = understaffing slack;

α_{il} = customer inconvenience cost per worker due to understaffing;

X_j = the number of workers working days off pattern j ;

C_j = cost of one worker on pattern j ;

A_{ij} = a 7 by J matrix where $a_{ij}=1$ if pattern j calls for working on day i and $a_{ij}=0$ otherwise

W_{total} = total size of the work force.

The objective of the model is to minimize the weekly labor cost of total workers plus the customer frustration cost due to understaffing. The first constraint of the model requires that the number of shifts X_j supplied by the days-off problem is equal to the number of shifts X_{ik} used in each of the shift scheduling problems. The second constraint requires that the personnel needed during each period of each day is equal

to the demand when overstaffing and understaffing are accounted for. The third constraint limits the shifts to the total available workers.

Bailey's model was for the 12-hour day, considering only 8-hour shifts and Ozkarahan used Bailey's model for limited staff size by extending the model to include 24-hour day and other shift combinations such as 8-hour, 10-hour and 12-hour shifts. For unlimited staff size Ozkarahan's model is based on the Field's (1983) model.

For unlimited staff size, it is necessary to solve the time-of-day(TOD) problems first and use their solutions as requirements for the day-of-week(DOW) scheduling problem.

TOD problem for unlimited staff size:

$$\min Z_i = \sum_{k=1}^{K_i} C_{ik} X_{ik}$$

subject to

$$\sum_{k=1}^{K_i} A_{ik} X_{ik} \geq D_{il} \quad l = 1, 2, \dots, L_i$$

$$X_{ik} \geq 0 \quad \text{and integer}$$

where

C_{ik} = cost of a nurse on day I and in pattern k;

i = Monday, Tuesday, ..., Sunday;

K_i = number of feasible pattern on day I;

L_i = number of periods to be scheduled on day I;

X_{ik} = number of nurses scheduled to work shift pattern k on day I;

D_{il} = demand for nurses during period l on day i;

A_{ik} = an L_i by K_i matrix where $a_{ik}=1$ if pattern k is worked during period l and

$a_{ik}=0$ otherwise

The TOD problem can use alternative objective functions that include measures of total cost, labor hours scheduled, staff size and percent idle time. The model is then solved for each day separately. And the sum of X's in each TOD solution represents the daily nursing requirements. Then the DOW model is:

$$\min Z = \sum_{j=1}^J C_j X_j$$

subject to

$$\sum_{j=1}^J A_{ij} X_j \geq D_i \quad i = 1, 2, \dots, 7$$

$$X_j \geq 0 \text{ and integer}$$

where

X_j = the number of nurses working days off pattern j;

C_j = cost of one nurse on pattern j;

A_{ij} = a 7 by J matrix where $a_{ij}=1$ if pattern j calls for working on day i and $a_{ij}=0$ otherwise

D_i = demand for nurses on day i;

J = number of feasible days off or work patterns.

If the staff size is limited, than DOW problem is solved first:

$$\min Z = \sum_{j=1}^J C_j X_j$$

subject to

$$\sum_{j=1}^J A_{ij} X_j \geq D_i \quad i = 1, 2, \dots, 7$$

$$\sum_{j=1}^J X_j = W_{total}$$

$$X_j \geq 0 \text{ and integer}$$

where

W_{total} = total work force and C_j , X_j , A_{ij} and D_i are as explained before.

To take a feasible solution, understaffing and overstaffing slacks are included in the model:

$$\min Z = \sum_{j=1}^J C_j X_j + \sum_{i=1}^7 \Gamma_i U_i$$

subject to

$$\sum_{j=1}^J A_{ij} X_{jk} - O_i + U_i = D_i \quad i = 1, 2, \dots, 7$$

$$\sum_{j=1}^J X_j = W_{total}$$

$$X_j \geq 0 \text{ and integer}$$

where

Γ_i = cost of understaffing by one nurse on day I;

O_i = level of overstaffing on day i;

U_i = level of understaffing on day i.

Constraints limiting the available nurses for each day need to be passed from the DOW problems solutions. Therefore TOD formulation takes the following form:

$$\min Z_i = \sum_{k=1}^{K_i} C_{ik} X_{ik}$$

subject to

$$\sum_{j=1}^J A_{ij} X_{ik} \geq D_{il} \quad l = 1, 2, \dots, L_i$$

$$\sum_{k=1}^{K_i} X_{ik} = W_i$$

$$\sum_{k \in 8N} X_{ik} = W_{8N}$$

$$\sum_{k \in 10N} X_{ik} = W_{10N}$$

$$\sum_{k \in 12N} X_{ik} = W_{12N}$$

$$\sum_{k \in 8D} X_{ik} = W_{8D}$$

$$\sum_{k \in 10D} X_{ik} = W_{10D}$$

$$\sum_{k \in 12D} X_{ik} = W_{12D}$$

$$\sum_{k \in 8E} X_{ik} = W_{8E}$$

$$\sum_{k \in 10E} X_{ik} = W_{10E}$$

$$\sum_{k \in 12E} X_{ik} = W_{12E}$$

$$X_{ik} \geq 0 \quad \text{and integer}$$

where

i = Monday, Tuesday, ..., Sunday;

K_i = number of feasible pattern on day i ;

L_i = number of periods to be scheduled on day i ;

W_i = sum of available supply of nurses on day i

$8N, 8D, 8E$ = set of indices for X_{ik} specifying the 8-hour shifts of the night, day and evening period segments of 24-hour day, respectively for the 10- and 12-hour shifts.

W_{8N}, W_{8D}, W_{8E} = workforce size for the 8-hour shifts of the night, day and evening period segments of 24-hour day, respectively for the 10- and 12- hour shifts.

For a feasible solution, understaffing and overstaffing slacks must be included in the model:

$$\min Z_i = \sum_{k=1}^{K_i} C_{ik} X_{ik} + \sum_{l=1}^{L_i} \alpha_{il} U_{il}$$

subject to

$$\sum_{k=1}^{K_i} A_{ik} X_{ik} - O_{il} + U_{il} = D_{il} \quad l = 1, 2, \dots, L_i$$

$$X_{ik} \geq 0 \quad \text{and integer}$$

where

i = Monday, Tuesday, ..., Sunday

α_{il} = cost of patient inconvenience per nurse due to understaffing during period l of day i ;

O_i = level of overstaffing during period l of day i ;

U_i = level of understaffing during period l of day i .

For solving the personnel scheduling problem, it is necessary to determine the requirements for personnel during each time period and queuing theory is used to calculate these requirements. An overview of queuing theory is given in the next section.

4.4 Queuing Theory

As it is mentioned above, queuing theory is used in this thesis to determine the number of personnel required during each time period. Then, these requirements are used as the right-hand side values for personnel scheduling models described in the previous section.

Queuing theory involves the mathematical study of waiting lines. In management science a *waiting line* is called a *queue*. There are a lot of books written on this subject. One of them Hillier & Lieberman, 1995 says that: "Waiting lines occur whenever the current demand for a service exceeds the current capacity to provide that service. Providing too much service involves excessive costs and not providing enough service capacity causes waiting line to become excessively long at times. Therefore, the ultimate goal is to achieve an economic balance between the cost of service and the cost associated with waiting for that service."

In other words, the usual objective of a queuing model is to determine how to provide service to customers in such a way that an efficient operation is achieved. Unlike the inventory or linear programming models, a minimum-cost or a maximum-profit solution is not always sought. Rather, the aim of these models is to determine

various characteristics of the queuing system, such as the mean waiting time and the mean length of the waiting line.

Queuing theory applies to service as well as manufacturing firms, relating customer arrival and service-system processing characteristics to service system output characteristics.

Table4.1 Common Queuing Situations

Situation	Arrivals in Queue	Service Process
Supermarket	Grocery shoppers	Checkout clerks at cash register
Highway toll booth	Automobiles	Collection of tolls at booth
Doctor's office	Patients	Treatment by doctors or nurses
Computer System	Programs to be run	Computer processes jobs
Telephone company	Callers	Switching equipment to forward calls
Bank	Customers	Transaction handled by teller
Machine maintenance	Broken machines	Repair people fix machines
Harbor	Ships and barges	Dock workers load and unload

4.4.1 Basic Structure Of Queuing Models

Analyzing queuing problems begins with a description of the situation's basic elements. Each specific situation has different characteristics, but four elements are common to all situations:

1. an input or customer population, that generates potential customers;
2. a waiting line of customers, a queue;
3. the service facility, consisting of a person, a machine or both necessary to perform the service for the customer;
4. a priority rule, which selects the next customer to be served by the service facility (Krajewski & Ritzman, 2002).

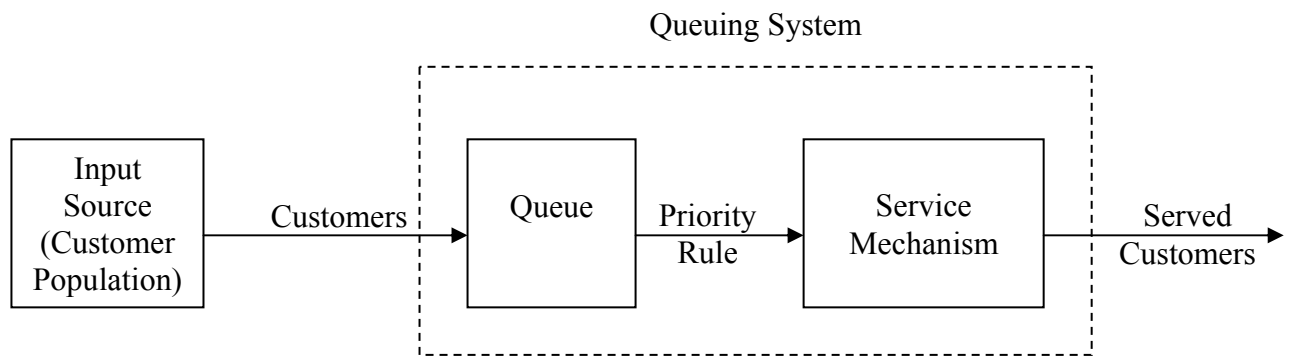


Figure4.1 The Basic Queuing Process

4.4.1 1 Customer Population

A customer population is the source of input to the service system. One characteristic of the input source is its size. The size is the total number of customers that might require service from time to time. This size may be assumed to be either finite or infinite. A finite population is a queue in which there are only a limited number of potential users of the service. Alternatively; infinite population is a queue in which a virtually unlimited number of people or items could request the services.

Customers arrive at a service facility either according to some known schedule, such as one patient every fifteen minutes, or else they arrive randomly. Frequently in queuing problems, the number of arrivals per unit time can be estimated by a probability distribution known as the Poisson distribution. An equivalent assumption is that the probability distribution of the time between consecutive arrivals refers to as the interarrival time, is an exponential distribution.

Customers in waiting lines may be patient or impatient. Patient customers are people or machines that wait in the queue until they are serve and do not switch between lines. Impatient customers are people who balk or renege. Customers who bulk refuse to join the waiting line because it is too long to suit their needs or interests. Reneging customers are those who enter the queue but then become impatient and leave without completing their transaction.

4.4.1.2 Queue

A queue is characterized by the maximum permissible number of customers that it can contain. Queues are called infinite or finite according to whether this number is infinite or finite (Hillier & Lieberman, 1995). Queues may be designed to be a single line or multiple lines. When multiple servers are available and each one can handle general transactions, the single line arrangement keeps servers uniformly busy and gives customers a sense of fairness (Krajewski & Ritzman, 2002).

4.4.1.3 Queue Discipline

The queue discipline refers to the order in which members of the queue are selected for service. It may be first-come-first-served (first-in-first-out FIFO), random, according to some priority procedure or some other order. Most systems use the queue discipline FIFO. A preemptive discipline is a rule that allows a customer of higher priority to interrupt the service of another customer.

4.4.1.4 Service Facility

Service facilities consist of the personnel and equipment necessary to perform the service for the customer and the service mechanism consists of one or more service facilities, each of which contains one or more parallel service channels, called servers. If there is more than one service facility, the customer may receive service from a sequence of these. Some services require a single step, also called phase, whereas others require a sequence of steps.

4.4.2 Probability Distributions

The sources of variation at service facilities in waiting-line problems come from the random arrivals of customers and the variations in service times. Each of these sources can be described with a probability distribution.

4.4.2.1 Arrival Distribution

Customers arrive at service facility either according to a schedule or randomly. The variability of customer arrivals often can be described by a Poisson distribution, which specifies the probability that n customers will arrive in T time periods:

$$P(n) = \frac{(\lambda T)^n}{n!} e^{-\lambda T} \quad \text{for } n = 0, 1, 2, \dots$$

where

λ = Arrival rate per period

T = Time period

n = Number of Arrivals

Another way to specify the arrival distribution is to do it in terms of customer interarrival times. If the customer population generates customers according to a Poisson distribution, the exponential distribution describes the probability that the next customer will arrive in the next T time periods.

In this thesis, the arrival rate of patients is assumed to be Poisson and therefore, interarrival times are assumed as exponentially distributed.

4.4.2.2 Service Time Distribution

Service patterns are like arrival patterns in that they may be constant or random. If service time is constant, it takes the same amount of time to take care of each customer. More often, service times are randomly distributed. The exponential distribution describes the probability that the service time of the customer at a particular facility will be more than T time periods:

$$P(t \leq T) = 1 - e^{-\mu T}$$

where

μ = mean number of customers completing service per period

t = service time of the customer

T = target service time.

The mean of the service time distribution is $1/\mu$ and the variance is $(1/\mu)^2$ (Krajewski & Ritzman, 2002).

In this thesis, there was no available data about service times and service time is assumed to be distributed exponentially with a mean rate that is determined according to the interviews.

4.4.3 Performance Measurement Of Queues

Queuing analysis can obtain many measures of a waiting line system's performance including the following:

4.4.3.1 The probability for the number of customers in the system

This distribution is the basis for establishing all the other results below. It may also be useful in designing facilities that can physically hold waiting customers.

4.4.3.2 The mean number of customers in the system

This quantity accounts for the number of customers either waiting in line or receiving service. It is useful primarily as an intermediate device for finding the mean customer time spent in the system. The number of customers in queue and being served also relates to service efficiency and capacity. A large number of customers in the system causes congestion and may result in customer dissatisfaction, unless more capacity is added.

4.4.3.3 The mean customer time spent in the system

This is an average amount representing the total time spent by a customer in the system. If some customers are spending too much time in the service system, there may be a need to change the priority discipline, increase productivity or adjust capacity in some way. When a cost may be associated with each unit of a customer's time, it may be used to make economic comparisons of alternative queuing systems.

4.4.3.4 Mean number of customers waiting (length of line)

This quantity is similar to the mean number of customers in the system but only, involves the customers who are actually waiting in line and not being serviced. The number of customers in the waiting line reflects one of two conditions. Short queues could mean either good customer service or too much capacity. Similarly, long queues could indicate either low server efficiency or the need to increase capacity. In this thesis, the primary performance measure of the management is the length of the queue. The management wants to keep the queue length within acceptable limits and doctor requirements are determined according to these limits. Then these requirements are used as the right-hand side values of the personnel scheduling model.

4.4.3.5 The mean customer waiting time

This is an average value that is used to evaluate the quality of service. Long lines do not always mean long waiting times. If the service rate is fast, a long line can be served efficiently. However, when waiting time seems long, customers perceive the quality of service to be poor. Managers may try to change arrival rate of customers or design the system to make long wait times seem shorter than they really are. For example, at Walt Disney World customers in line for an attraction are entertained by videos and also are informed about expected waiting times, which seems to help them endure the wait.

4.4.3.6 The server utilization factor

This is the proportion of time that the server actually spends with customers, the time during which server is busy. It provides an estimate of the expected amount of server idle time that may be devoted to secondary tasks not directly involved with service.

4.4.4 Basic Queuing Systems

Service systems are usually classified in terms of their number of channels and number of phases.

In the ***single-channel, single-phase queuing system***, all services demanded by a customer can be performed by a single-server facility. Customers form a single line and go through the service facility one at a time. Examples are a drive-through car wash and a machine that must process several batches of parts.

The single-channel, multiple-phase arrangement is used when the services are best performed in sequence by more than one facility, yet customer volume or other constraints limit the design to one channel. Customers form a single line and proceed sequentially from one service facility to the next. An example of this arrangement is a McDonald's drive through, where the first facility takes the order, the second takes the money and the third provides the food.

The multiple-channel, single-phase arrangement is used when demand is large enough to warrant providing the same service at more than one facility or when the services offered by the facilities are different. Customers form one or more lines, depending on the design. In the single-line design, customers are served by the first available server, as in the lobby of a bank. If each channel has its own waiting line, customers wait until the server for their line can serve them, as at a bank's drive-through facilities.

The multiple-channel, multiple-phase arrangement occurs when customers can be served by one of the first-phase facilities but then require service from a second-phase facility, and so on. In some cases, customers cannot switch channels after service has begun; in others they can. An example of this arrangement is a Laundromat. Washing machines are the first-phase facilities and dryers are the second-phase facilities. Some of the washing machines and dryers may be designed for extra-large loads, thereby providing the customer a choice of channels.

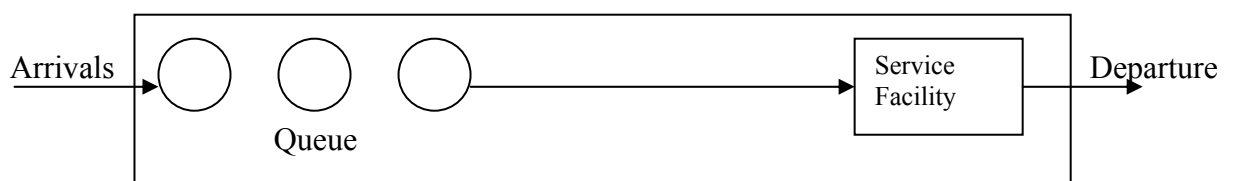


Figure 4.2 Single-channel, single-phase system

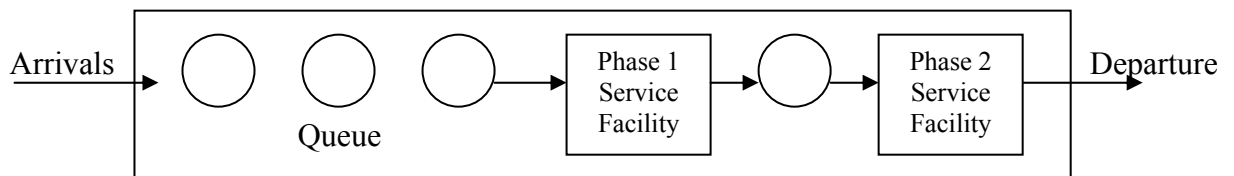


Figure 4.3 Single-channel, multiple-phase system

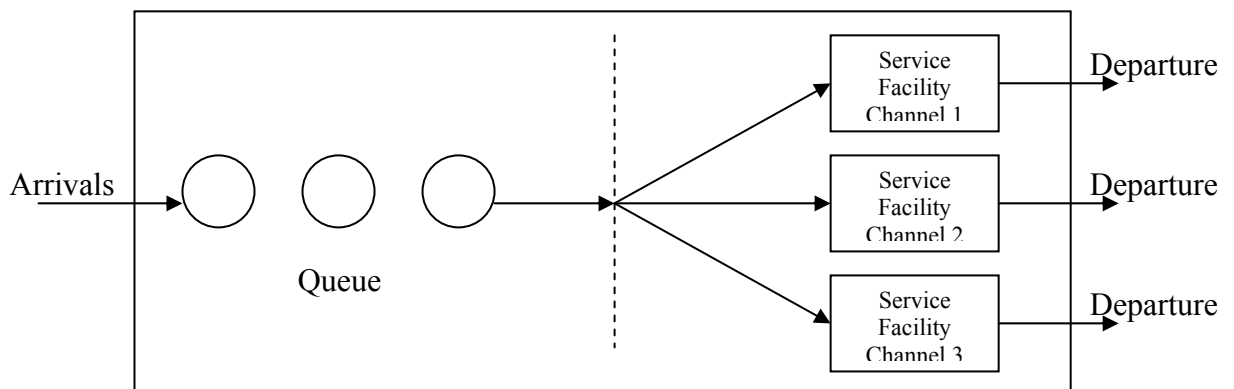


Figure 4.4 Multiple-channel, single-phase system

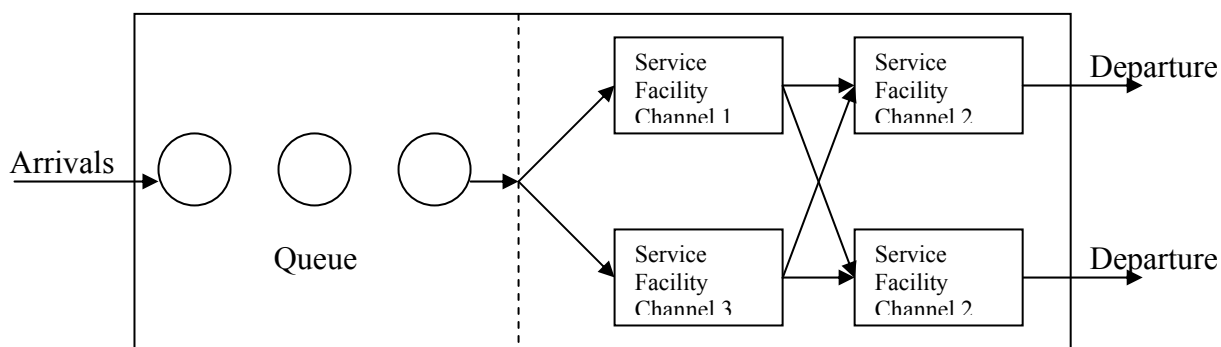


Figure 4.5 Multiple-channel, multiple-phase system

4.4.5 Variety Of Queuing Models

Many of the queuing models assume that all interarrival times are independent and identically distributed and that all service times are independent and identically distributed. Such models conventionally are labeled as follows:

Distribution of interarrival times / Distribution of service times / Number of servers

where

M = Exponential distribution (Markovian)

D = Degenerate distribution (constant times)

E_k = Erlang distribution

G = General distribution (any arbitrary distribution allowed)

4.4.5.1 The Single-Channel Queuing Model (M / M / 1)

The most common case of queuing problems involves the single-channel waiting line. In this situation, arrivals form a single line to be serviced by a single station. It is assumed that the following conditions exist in this type of system:

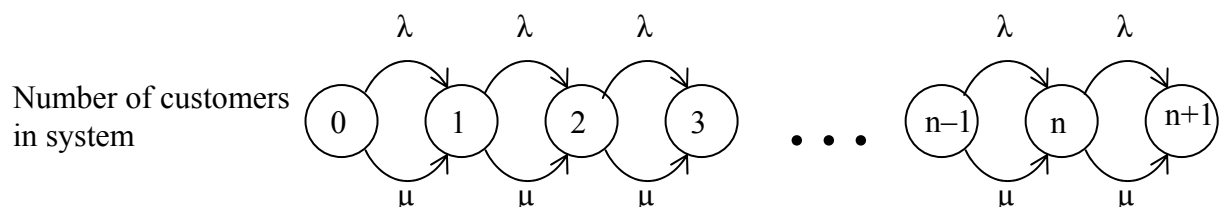
- i) Arrivals are served on a first-in-first-out (FIFO) basis and every arrival waits to be served regardless of the length of the queue.
- ii) Arrivals are independent of preceding arrivals but the average number of arrivals (arrival rate) does not change over time.
- iii) Arrivals are described by a Poisson distribution and come from an infinite population. Thus, the interarrival times have an exponential distribution and the number of customers arriving in any specified time interval has a Poisson distribution. Arrivals occur at a **mean arrival rate λ** .
- iv) Service times vary from one customer to the next and are independent of one another but their average rate is known.
- v) Service times are exponentially distributed and the **mean service rate** is represented by μ . Thus, the mean service time is $1/\mu$.
- vi) The mean service rate must exceed the mean arrival rate, $\mu > \lambda$, so that the queuing system must have more than enough capacity to service all customers. Without this restriction, a queuing system would be unstable and the waiting line would grow indefinitely.

Basic Queuing Formulas:

The formulas are in terms of two parameters:

λ = Mean customer arrival rate

μ = Mean service rate



- *Probability that no customers are in the system (the service unit is idle)*

$$P_0 = 1 - \frac{\lambda}{\mu}$$

- *Probability of n units in the system*

$$P_n = \left(\frac{\lambda}{\mu} \right)^n P_0$$

- *Mean number of customers in the system:* λ may be viewed as a measurement of the “demand” for service and μ as an expression of the “capacity” of the service facility. The difference $\mu - \lambda$ then represents the “excess capacity” of the system to fill demand. Thus, the mean number of customers in the system is the ratio of demand to excess capacity.

$$L_s = \frac{\lambda}{\mu - \lambda}$$

- *Mean customer time spent in the system:* The mean customer time spent in the system is determined by multiplying the mean time between customer arrivals $1/\lambda$ and the mean number of customers L .

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu - \lambda}$$

- *Mean number of customers waiting (length of line)*

$$L_q = L_s - \frac{\lambda}{\mu} = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

- *Mean customer waiting time:* A customer's mean waiting time is simply the difference between the mean time that customer spends in the system and the mean service time.

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu - \lambda)} = W_q - \frac{1}{\mu}$$

- *Server utilization factor:*

$$\rho = \frac{\lambda}{\mu}$$

4.4.5.2 The Multiple-Channel Queuing Model (M / M / s)

The system analyzed in this thesis a multiple-channel system and queuing formulas related to this variety are used to calculate the performance measures of the system. It is assumed that the following conditions exist in this type of system:

- 1) Arrivals are served on a first-in-first-out (FIFO) basis and every arrival waits to be served regardless of the length of the queue.
- 2) Arrivals are independent of preceding arrivals but the average number of arrivals (arrival rate) does not change over time.
- 3) Arrivals are described by a Poisson distribution and come from an infinite population. Thus, the interarrival times have an exponential distribution and the number of customers arriving in any specified time interval has a Poisson distribution. Arrivals occur at a **mean arrival rate λ** .
- 4) Service times vary from one customer to the next and are independent of one another but their average rate is known.
- 5) Service times are exponentially distributed and the **mean service rate** is represented by **μ** . Thus, the mean service time is **$1/\mu$** .
- 6) The mean service rate must exceed the mean arrival rate, **$\mu > \lambda$** , so that the queuing system must have more than enough capacity to service all customers. Without this restriction, a queuing system would be unstable and the waiting line would grow indefinitely.

- 7) Each service facility is identical in all respects and that each is capable of performing service at the rate of μ customers per unit time.

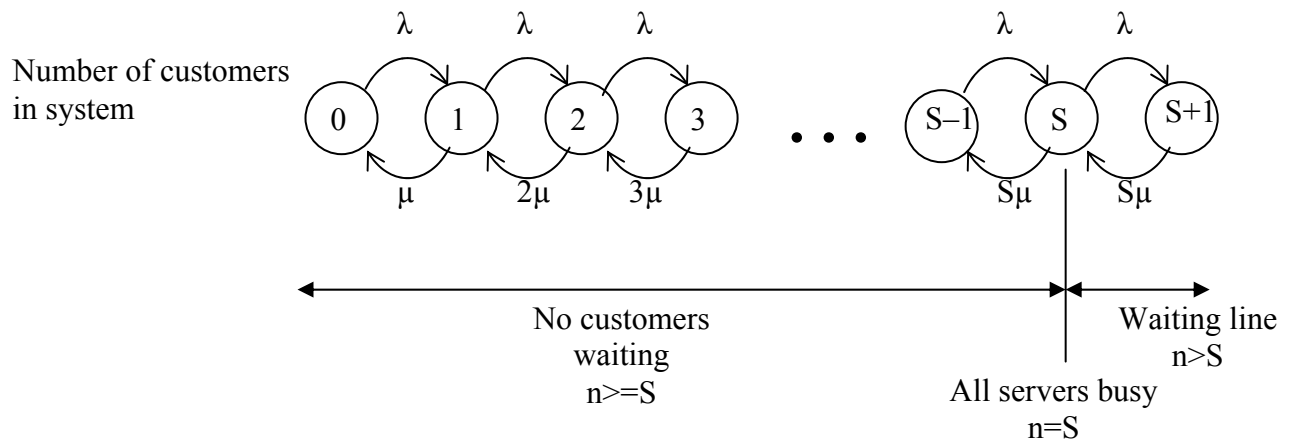
Basic Queuing Formulas:

The formulas are in terms of two parameters:

λ = Mean customer arrival rate

μ = Mean service rate

S = Number of service channels



- Probability of that no customers are in the system (the service unit is idle)

$$P_0 = \frac{1}{\left[\sum_{n=0}^{s-1} \frac{(\lambda / \mu)^n}{n!} + \frac{(\lambda / \mu)^s}{s!} \left(\frac{1}{1 - \lambda / s\mu} \right) \right]}$$

- Probability that n customers are in the system

$$P_n = \begin{cases} \frac{(\lambda / \mu)^n}{n!} P_0 & \text{if } 0 \leq n \leq s \\ \frac{(\lambda / \mu)^n}{s! s^{n-s}} P_0 & \text{if } n \geq s \end{cases}$$

- Mean number of customers waiting (length of line)

$$L_q = \frac{(\lambda / \mu)^s (\lambda / s\mu)}{s! (1 - \lambda / s\mu)} P_0$$

- *Mean customer waiting time*

$$W_q = \frac{L_q}{\lambda}$$

- *Mean customer time spent in the system*

$$W_s = W_q + \frac{1}{\mu}$$

- *Mean number of customers in the system*

$$L_s = L_q + \frac{\lambda}{\mu}$$

- *Server Utilization Factor*

$$\rho = \frac{\lambda}{s\mu}$$

4.4.5.3 The Single-Channel Queuing Model for a Finite Queue

The number of customers that the queuing system can handle at any given time is limited. A waiting line of limited length is called a finite queue. A finite queue may arise due to a physical constraint or customers may simply give up when the waiting line becomes too long. In either case, a customer who is turned away does not return to the system.

Basic Queuing Formulas:

The formulas are in terms of these parameters:

λ = Mean customer arrival rate

μ = Mean service rate

M = Maximum number of customers in the system

- *Probability of that no customers are in the system (the service unit is idle)*

$$P_0 = \frac{(1 - \lambda / \mu)}{1 - (\lambda / \mu)^{M+1}}$$

- *Probability that n customers are in the system*

$$P_n = (\lambda / \mu)^n P_0 \quad \text{for } 1 \leq n \leq M$$

- *Mean number of customers in the system*

$$L_s = \frac{\lambda / \mu}{1 - \lambda / \mu} - \frac{(M + 1)(\lambda / \mu)^{M+1}}{1 - (\lambda / \mu)^{M+1}}$$

- *Mean customer time spent in the system*

$$W_s = \frac{L}{\lambda(1 - P_M)}$$

- *Mean number of customers waiting (length of line)*

$$L_q = L_s - (1 - P_0)$$

- *Mean customer waiting time*

$$W_q = \frac{L_q}{\lambda(1 - P_M)}$$

4.4.5.4 The Single-Channel Queuing Model with Constant Service Times (M / D / 1)

Service times are constant and the **mean service rate** is represented by μ . Thus, the mean service time is $1/\mu$. Because constant rates are certain the values for L_q , W_q , L_s and W_s are always less than they would be in models with exponentially distributed service times because of variable service rates.

Basic Queuing Formulas:

The formulas are in terms of these parameters:

λ = Mean customer arrival rate

μ = Mean service rate

M= Maximum number of customers in the system

- *Mean number of customers waiting (length of line)*

$$L_q = \frac{\lambda^2}{2\mu(\mu - \lambda)}$$

- *Mean customer waiting time*

$$W_q = \frac{\lambda}{2\mu(\mu - \lambda)}$$

- *Mean number of customers in the system*

$$L_s = L_q + \frac{\lambda}{\mu}$$

- *Mean customer time spent in the system*

$$W_s = W_q + \frac{1}{\mu}$$

4.4.5.5 The Single-Channel Queuing Model for a Limited Population

Customers arriving at the system represent a small population. Although this case resembles the model for a finite queue, here the potential customers are limited. The model differs from the earlier models because there is now a dependent relationship between the length of queue and the arrival rate.

Basic Queuing Formulas:

The formulas are in terms of these parameters:

λ = Mean customer arrival rate

μ = Mean service rate

M= Maximum number of customers that may need system

- *The probability for the number of customers either waiting or receiving service:*

$$P_0 = \frac{1}{\sum_{n=0}^M \left[\frac{M!}{(M-n)!} \left(\frac{\lambda}{\mu} \right)^n \right]}$$

$$P_n = \frac{M!}{(M-n)!} \left(\frac{\lambda}{\mu} \right)^n P_0 \quad \text{for } 1 \leq n \leq M$$

- *Mean number of customers waiting (length of line)*

$$L_q = M - \frac{\lambda + \mu}{\lambda} (1 - P_0)$$

- *Mean number of customers in the system*

$$L_s = L_q + (1 - P_0)$$

- *Mean customer waiting time*

$$W_q = \frac{L_q}{\lambda(M - L_s)}$$

- *Mean customer time spent in the system*

$$W_s = \frac{L_s}{\lambda(M - L_s)}$$

CHAPTER FIVE
APPLICATION IN DOKUZ EYLUL UNIVERSITY
DEPARTMENT OF ORTHOPEDICS AND TRAUMATOLOGY

5.1 Introduction

The Department of Orthopedics and Traumatology started receiving and treating patients in 1978 at the Eşrefpaşa Hospital as a clinic of the Izmir Faculty of Medicine of the Ege University and later, upon the establishment of the Dokuz Eylul University in 1982, it became the Department of Orthopedics and Traumatology of this new university.

Today research, medical care and education programs are undertaken by a surgical and academic team consisting of 8 professors, 5 associate professors, 2 assistant professors, 1 lecturer and 24 residents. The outpatient clinic is served by research assistants and the rest of the team serves patients by appointment.

More than 100 patients in a day are seen in outpatient clinics. The Department of Orthopedic and Traumatology has 52 beds and 17 operating theatre rooms for week, which allow them to perform 50 surgeries in a week.

There are some subspecialties in the department to gather and excel in the knowledge and research.

Spinal Surgery Unit

The spinal surgery team performs spinal surgeries of all kinds including congenital or acquired spinal deformities, spinal degeneration and related problems, and traumatic conditions. The spinal unit is a main referral centre is given credit to be one of the nation's leading centers for spinal problems.

Hand and Upper Extremity Surgery Division

The division of Hand and Upper Extremity Surgery was instituted in the Department of Orthopedic and Traumatology in 1993. All kinds of deformities and traumas in hand and upper extremity are taken care of by this division. The division also performs major reimplantations and microsurgeries.

Arthroplasty

This unit provides superior care for patients with joint problems by performing arthroplasty in accordance with the latest developments.

Arthroscopy, Sports Traumatology and Knee Surgery

Treatment or repair of meniscus lesions, repairs of anterior and posterior cruciate ligament, arthroscopy of knee, ankle and hip for various lesions, sports traumatology are among the subjects of main interest of unit.

Oncology

This unit provides orthopedic treatment of all musculoskeletal tumors using various methods including limb salvage methods.

Research and General Orthopedics

This unit deals with research and scaphoid fracture.

Pediatric Orthopedics

Orthopedic problems in children must be evaluated separately from adult patients. This unit provides medical and surgical care for children challenged by neuromuscular disorders such as cerebral palsy, deformities and congenital or developmental anomalies such as club-foot, dysplasia of the hip.

Professors, associate professors and assistant professors treat patients on an appointment based system so the first step of the study is to schedule appointments of doctors by using dynamic programming approach. Because there is not an available tool to solve the dynamic programming problem, calculations are made manually so appointments of only one doctor are scheduled. The others can be calculated similarly. Since, decisions cannot be made prior to the clinic session; dynamic programming is not an efficient approach for real-time problems. Then, the problem is solved by using branch-and-bound approach for the static state in which decisions can be made prior to the clinic session. Calculations for the dynamic programming are used to obtain static state solution.

The next step of the study is scheduling work hours of personnel treating patients in the outpatient clinic. Because doctors' have some other duties than treating patients, they are not able to work through the whole clinic session and shifts are scheduled in order to provide leisure time for doctors. Number of doctors required for each time period is determined by using queuing theory approach. The first step of the queuing theory is analyzing the data about patient arrivals in order to determine the patient arrival rate for each time period of the day. The data is gathered from Department of Orthopedics and Traumatology patient records for 23 weeks. The time periods last one hour and there are eight periods per day. After obtaining patient arrival rates and doctor service rate, the system is analyzed by using queuing formulas and doctor requirements for each period and each day are determined to keep patient waiting time within a limit. Then by taking into consideration these requirements, available shifts for doctors are determined and a personnel scheduling model is built in order to meet patient demand. A linear program is developed for shift scheduling to cover the doctor requirements per period and per day, then by using a simple assignment heuristic, doctors are assigned to these shifts.

5.2 Appointment Scheduling Problem

In the department, there are four professors, six associate professors and one assistant professor that treat patients on an appointment basis. The appointment intervals and number of patients to be scheduled changes doctor by doctor but generally appointments are given by an interval of fifteen minutes. In other words four patients are seen per hour and it means that the current appointment rule used is individual-block fixed-interval rule.

The patient flow is very simple that once the patient is treated by the doctor he/she exits. The patients take the appointment via the secretary of the doctor and it is not necessary to make another registration before the consultation. So, the system is a single-phase system and appointments are scheduled for each doctor separately. In other words, there is only one doctor for each scheduling problem.

The proposed system in this study is a variable-sized multiple-block appointment system with only one server. In this system patients are scheduled at the start of each period. Here, a time period refers to an hourly interval. The objective of the model is calculating how many patients will be scheduled to each block in order to minimize patient waiting time and doctor's idle time.

The assumptions of the study are as follows:

- 1) Patients are punctual.
- 2) The no-show probability is 0.
- 3) Walk-ins and emergencies are not taken into consideration.
- 4) Doctor starts the clinic on time.
- 5) During consultation, there is no interruption for doctor.
- 6) The patients are treated on a first-come first-served basis.

As it is described in Chapter three, the formulations of Liao, Pegden and Rosenshine (1993) is used and modified in the study. According to their formulation the general form of the problem is as follows:

$$\begin{aligned}
 f_i(r, s) &= \min_{0 \leq d_i \leq r} g_i(r, s, d_i) \\
 &= \min_{0 \leq d_i \leq r} \left\{ \frac{c_w d_i}{2\mu} (2s + d_i - 1) + \left[\sum_{j=0}^{s+d_i} P(s + d_i - j | s + d_i) * f_{i+1}(r - d_i, s + d_i - j) \right] \right\} \\
 &\quad i = 1, 2, \dots, k
 \end{aligned} \tag{1}$$

The first term of the formulation indicates waiting cost of the patients while the second refers to the cost of the next period. As it seen, there is no term about idle time in the formulation. Since the objective of study is minimizing patient waiting time and doctor idle time, it is necessary to take into consideration the cost of idle time.

Given that s patients remain in the system from the previous periods, if d_i patients are assigned to the period, j of $(s+d_i)$ patients will be treated with a probability of $p(j)$. The average consultation time of each patient is defined in terms of the time period. In this study; a time period is one time unit while average consultation time is 0,3944 time units. Then, the expected idle time of the doctor can be calculated as follows:

$$c_I \sum_{j=0}^{s+d_i} (p(j) * (1 - j\mu))$$

where

c_I : doctor idle time cost per unit time.

S, d_i, j and μ are as described in Chapter three.

Then, the modified formulation of the problem is as follows:

$$\begin{aligned}
f_i(r, s) &= \min_{0 \leq d_i \leq r} g_i(r, s, d_i) \\
&= \min_{0 \leq d_i \leq r} \left\{ \frac{c_w d_i}{2\mu} (2s + d_i - 1) + \left[\sum_{j=0}^{s+d_i} P(s + d_i - j \mid s + d_i) * (c_l * (1 - j * \mu) + f_{i+1}(r - d_i, s + d_i - j)) \right] \right\} \\
&\quad i = 1, 2, \dots, k
\end{aligned} \tag{2}$$

However, the boundary condition of the model does not change. The clinic session finishes when all patients are treated by the doctor, so there is no idle time for the last period. The doctor can leave the clinic when all patients are seen whether it is the last period or not.

$$\begin{aligned}
f_k(r, s) &= g_k(r, s, r) \\
&= \frac{c_w r}{2\mu} (2s + r - 1) + c_s \sum_{j=0}^{r+s-1} P(r + s - j \mid r + s) \frac{r + s - j}{\mu}
\end{aligned} \tag{3}$$

5.2.1 Model Parameters

The first parameter to decide is the number of time periods per day. As it is mentioned above scheduling of appointments is done for only one doctor. Appointment period of the doctor changes day to day because of some other duties of the doctor. For example some days, he treats patients for three time periods while it is four on other days so two alternative models are solved by considering three time periods and four time periods respectively ($k=2$ and $k=3$).

Also, it is necessary to determine the total number of patients to be scheduled during a clinic session. The historical data about patient treatment between 10.01.2005 and 27.05.2005 is analyzed for this doctor and Table 4.1 shows the number of patients treated per hour. It is seen that totally 431 patients are seen during 84 days. Some days there was no available record because no patients were treated because of other duties of the doctor. By making a simple calculation, it is found that $431/84=5,131$ patients are seen per day so it is necessary to schedule approximately six patients per day. This means that, for the first period, the r value, that shows the number of patients unscheduled, is equal to six given that there is no patient remaining in the system from previous periods, $s=0$.

The third parameter to be determined is the probabilities for consultation completions. The data about patient treatments between 10.01.2005 and 27.05.2005 is also analyzed for determining these inputs. The data on the number of patients seen per hour is seen on Table 5.1 and Table 5.2 shows probabilities of consultation completions per hour. As it seen, the highest probability of consultation completion with 0,25882 shows that most probably two patients can be seen during a time period.

Table 5.1 Consultation Completions per Hour

4	1	3	4	4
3	4	2	2	4
4	2	1	4	4
5	1	2	1	1
5	1	3	3	3
2	3	1	1	4
3	4	3	1	2
4	2	3	3	1
2	2	3	4	3
4	2	4	1	2
4	3	2	4	4
3	1	3	4	4
4	3	3	2	3
1	1	3	2	1
3	2	1	2	4
1	2	2	3	1
5	1	2	3	3
2	1	3	4	1
2	2	4	1	3
3	3	2	4	4
2	1	1	2	4
3	4	3	2	2
1	2	1	2	4
3	1	3	2	4
1	3	4	3	2
2	2	2	1	1
4	2	2	2	3
1	2	3	5	2
1	3	3	1	3
2	2	3	5	2
1	1	3	2	1
1	1	3	5	1
6	1	2	4	3
4	3	1	4	1

Table 5.2 Probabilities of Consultation Completions per hour

Consultation Completion per Hour	Frequency	Probability
1	42	0,247058824
2	44	0,258823529
3	43	0,252941176
4	34	0,2
5	6	0,035294118
6	1	0,005882353
TOTAL	170	1,00

However, these probabilities depend on the number of patients assigned to the period. According to the data, none of the consultations lasts longer than a time period so if only one patient is present in the system, then consultation completion probability of this patient will be 1,00. This situation exists if there is one patient remaining in the system from the previous periods and no patient is assigned to the present period or one patient is assigned to the present period while there is no patient remaining from previous periods. In such cases, the doctor will be idle for the rest of the time period. As a result, these probabilities are valid when two or more patients are assigned to a period.

The next parameter to be determined is about the consultation times. A time period is considered to be one time unit so 60 minutes is equal to one time unit. Then, the consultation times are defined in terms of time units. The average consultation time can be determined as thirty minutes, in other words 0,5 time units according to the interviews with the personnel but it can also be determined according to these probabilities as in Table 5.3. Since an average of 2,535 consultations are completed during a time period, average consultation time can be calculated simply by dividing a time period, 60 min, by 2,535. Then, the average consultation time can be found as 23,666 minutes or in other words 0,394 time units. As it is mentioned above all consultations last shorter than a hour but consultation times can vary between 10 minutes and 40 minutes.

Table 5.3 Average Consultation Time

Consultation Completion per Hour	Probability	
1	0,247058824	0,247058824
2	0,258823529	0,517647059
3	0,252941176	0,758823529
4	0,2	0,8
5	0,035294118	0,176470588
6	0,005882353	0,035294118
TOTAL	1,00	2,535294118
Average Consultation Time (min)		23,66589327
Average Consultation Time (time units)		0,394431555

The last parameters of the model are about costs. Since it is too difficult to express waiting cost, idle time cost as monetary values, it is better to use a proportion for these costs. Generally, patient waiting cost has the least weight for health care systems. Doctor idle time and over time cost is related to the hourly wage of the doctor so they have the same value.

In this study, $c_w / c_s / c_l$ costs are proportional to 1 / 3 / 3. This means that idle time and overtime cost of the doctor are three times bigger than the cost of patient waiting. First, solution for the dynamic state is obtained. Then this solution is used for solving the static state.

5.2.2 Dynamic State

Two alternative dynamic programming models are constructed with two different numbers of periods for the dynamic state. In the first alternative model the number of periods (k) is four and in the second alternative k is three.

The parameters of the model calculated in the previous section are as follows:

Number of patients require service (N) = 6;

Patient waiting cost per unit time (c_w) = 1;

Doctor idle time cost per unit time (c_l) = 3;

Doctor overtime cost per unit time (c_s) = 3;

Mean service time (μ) = 0,3944 time unit;

Probabilities of service completions ($p(j)$):

$$p(0) = 0$$

$$p(1) = 0,247058824$$

$$p(2) = 0,258823529$$

$$p(3) = 0,252941176$$

$$p(4) = 0,2$$

$$p(5) = 0,035294118$$

$$p(6) = 0,005882353$$

Two alternative models are proposed for the appointments of the doctor. The first alternative model has four time periods and the second alternative has three time periods during a clinic session.

5.2.2.1 Alternative Model 1

The first model has four time periods to schedule patients and according to the model parameters determined in the previous section, the formulations given above are used to construct the f-table set. As it described in Chapter three, calculations start from the last (4th) period.

- *Time Period4*

For the last period, all remaining patients must be scheduled so the optimum decision variable $P(r,s)$ is equal to the number of remaining patients unscheduled (r). Also overtime for doctor is possible while idle time is ignored for this period. The costs of the last period are calculated by using formula (3) and the results are seen on Table 5.4.

Table 5.4 f table set for the last period

r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	1,88	0	5,73	0	11,5	0	18,9	0	26,4	0
1	0	1	4,41	1	10,8	1	19,2	1	29	1	39,1	1		
2	4,41	2	13,3	2	24,2	2	36,6	2	49,3	2				
3	13,3	3	26,8	3	41,7	3	56,9	3						
4	26,8	4	44,2	4	61,9	4								
5	44,2	5	64,5	5										
6	64,5	6												

The r and s values change in the range from zero to six. Since there are six patients to be scheduled the sum of r and s cannot exceed six. $f(r,s)$ indicates cost of the state (r,s) for the last period and $P(r,s)$ indicates the number of patients scheduled in the last period. These values are then used for the calculation of other periods.

- *Time Period 3*

For the 3rd period, after calculating the values for $s=0$, the theorems described in Chapter three can be used for further calculations.

For $r=0$ and $s=0$, it is obvious that both of $f(0,0)$ and $P(0,0)$ values will be zero.

To determine the other values for $s=0$, less calculations can be made by using equation (8) given in Chapter three. This equation limits the d_i values possible.

$$f_i(r,0) = \min_{P_i(r-1,0) \leq d_i \leq \min\{P_i(r-1,0)+1, P_{i+1}(r,0)\}} g_i(r,0, d_i) \quad i \leq k-1, r \geq 1$$

According to this formulation the bounds for the d value are as:

$$P_i(r-1,0) \leq d_i \leq \min\{P_i(r-1,0)+1, P_{i+1}(r,0)\}$$

The bounds for all combinations of r and s , given that $s=0$, values and corresponding $f(r,s)$ values are as follows:

➤ For $s=0$ and $r=1$:

$$P_3(0,0) \leq d_i \leq \min \{P_3(0,0)+1, P_4(1,0)\}$$

$$0 \leq d_i \leq \min (1,1) = 0 \leq d_i \leq 1$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	1	$0 \leq d \leq 1$	0	3	1
			1	0	

➤ For $s=0$ and $r=2$:

$$P_3(1,0) \leq d_i \leq \min \{P_3(1,0)+1, P_4(2,0)\}$$

$$1 \leq d_i \leq \min (2,2) = 1 \leq d_i \leq 2$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	2	$1 \leq d \leq 2$	1	1,8168	1
			2	3,148344	

➤ For $s=0$ and $r=3$:

$$P_3(2,0) \leq d_i \leq \min \{P_3(2,0)+1, P_4(3,0)\}$$

$$1 \leq d_i \leq \min (2,3) = 1 \leq d_i \leq 2$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	3	$1 \leq d \leq 2$	1	6,231548	2
			2	4,239046	

➤ For $s=0$ and $r=4$:

$$P_3(3,0) \leq d_i \leq \min \{P_3(3,0)+1, P_4(4,0)\}$$

$$2 \leq d_i \leq \min (3,4) = 2 \leq d_i \leq 3$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	4	$2 \leq d \leq 3$	2	7,5852	2
			3	12,02978	

➤ For $s=0$ and $r=5$:

$$P_3(4,0) \leq d_i \leq \min \{P_3(4,0)+1, P_4(5,0)\}$$

$$2 \leq d_i \leq \min (3,5) = 2 \leq d_i \leq 3$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	5	$2 \leq d \leq 3$	2	10,67447	2
			3	18,79719	

➤ For $s=0$ and $r=6$:

$$P_3(5,0) \leq d_i \leq \min \{P_3(5,0)+1, P_4(6,0)\}$$

$$2 \leq d_i \leq \min (3,5) = 2 \leq d_i \leq 3$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	6	$2 \leq d \leq 3$	2	21,0028	2
			3	28,89847	

The $f(r,s)$ and $P(r,s)$ values for $s=0$ in the second period are as in Table 5.5:

Table 5.5 $f(r,s)$ and $P(r,s)$ values for $s=0$ in the 3rd period

s	r	d	f (r,s)	P (r,s)
0	0	0	0	0
0	1	0	3	1
		1	0	
0	2	1	1,8168	1
		2	3,14834402	
0	3	1	6,23154764	2
		2	4,23904638	
0	4	2	7,58520049	2
		3	12,0297775	
0	5	2	10,674471	2
		3	18,7971856	
0	6	2	21,0028047	2
		3	28,8984651	

Theorem 2 shows that if there are no patients remaining in the system from previous periods and there are patients unscheduled, then the number of patients scheduled to that period cannot be zero. So, the $P(r,s)$ values must be at least one except state (0,0). By selecting the d_i values with minimum cost $P(r,s)$ values are determined as in the last column.

By using Theorem 5, the $P(r,s)$ values can be determined for all other s and r combinations and $f(r,s)$ values are calculated for only these $P(r,s)$ values.

$$P_i(r,s) = (P_i(r+1,s-1) - 1)^+ \text{ for all } i, r \leq N-1 \text{ and } s \geq 1$$

The right side of the formulation indicates that, if the value of $P_i(r+1,s-1)-1$ is less than zero, then $P_i(r,s)$ is equal to zero. Then, the necessary calculations are as follows:

Table 5.6 $f(r,s)$ and $P(r,s)$ values in the 3rd period

s	r	P(r+1,s-1)	P(r,s)	f(r,s)
1	0	1	0	0
	1	1	0	1,8168
	2	2	1	4,239046
	3	2	1	9,82706
	4	2	1	13,20997
	5	2	1	21,0028
2	0	0	0	0,448856
	1	1	0	0,612847
	2	1	0	5,049704
	3	1	0	10,67447
	4	1	0	18,46731
3	0	0	0	1,077133
	1	0	0	4,423287
	2	0	0	11,19069
	3	0	0	21,29197
4	0	0	0	2,514207
	1	0	0	9,281615
	2	0	0	20,26584
5	0	0	0	5,433625
	1	0	0	16,41785
6	0	0	0	10,12386

Then the f -set table for the 3rd period takes the following form:

Table 5.7 f table set for the 3rd period

r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	0,45	0	1,08	0	2,51	0	5,43	0	10,1	0
1	0	1	1,82	0	0,61	0	4,42	0	9,28	0	16,4	0		
2	1,82	1	4,24	1	5,05	0	11,2	0	20,3	0				
3	4,24	2	9,83	1	10,7	0	21,3	0						
4	7,59	2	13,2	1	18,5	0								
5	10,7	2	21	1										
6	21	2												

- *Time Period 2*

Now, the same calculations are made for the second period as in the third period. For $r=0$ and $s=0$, both of $f(0,0)$ and $P(0,0)$ values will be zero. The other values for $s=0$ are as follows:

➤ For $s=0$ and $r=1$:

$$P_2(0,0) \leq d_i \leq \min\{P_2(0,0)+1, P_3(1,0)\}$$

$$0 \leq d_i \leq \min(1,1) = 0 \leq d_i \leq 1$$

s	r	Bounds	d	f(r,s)	P(r,s)
0	1	$0 \leq d \leq 1$	0	0	1
			1	0	

➤ For $s=0$ and $r=2$:

$$P_2(1,0) \leq d_i \leq \min\{P_2(1,0)+1, P_3(2,0)\}$$

$$1 \leq d_i \leq \min(2,1) = 1 \leq d_i \leq 1$$

s	r	Bounds	d	f(r,s)	P(r,s)
0	2	$1 \leq d \leq 1$	1	1,8168	1

➤ For $s=0$ and $r=3$:

$$P_2(2,0) \leq d_i \leq \min \{P_2(2,0)+1, P_3(3,0)\}$$

$$1 \leq d_i \leq \min (2,2) = 1 \leq d_i \leq 2$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	3	$1 \leq d \leq 2$	1	3,6336	2
			2	3,597200	

➤ For $s=0$ and $r=4$:

$$P_2(3,0) \leq d_i \leq \min \{P_2(3,0)+1, P_3(4,0)\}$$

$$1 \leq d_i \leq \min (2,2) = 1 \leq d_i \leq 2$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	4	$1 \leq d \leq 2$	1	6,055846	2
			2	4,665868	

➤ For $s=0$ and $r=5$:

$$P_2(4,0) \leq d_i \leq \min \{P_2(4,0)+1, P_3(5,0)\}$$

$$2 \leq d_i \leq \min (3,2) = 2 \leq d_i \leq 2$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	5	$2 \leq d \leq 2$	2	6,673371	2

➤ For $s=0$ and $r=6$:

$$P_2(5,0) \leq d_i \leq \min \{P_2(5,0)+1, P_3(6,0)\}$$

$$2 \leq d_i \leq \min (3,2) = 2 \leq d_i \leq 2$$

s	r	Bounds	d	f (r,s)	P (r,s)
0	6	$2 \leq d \leq 2$	2	8,375212	2

The $f(r,s)$ and $P(r,s)$ values for $s=0$ in the second period are as in Table 5.8:

Table 5.8 $f(r,s)$ and $P(r,s)$ values for $s=0$ in the 2nd period

s	r	d	f (r,s)	P (r,s)
0	0	0	0	0
0	1	0	0	1
		1	0	
0	2	1	1,8168	1
0	3	1	3,6336	2
		2	3,597200	
0	4	1	6,055846	2
		2	4,665868	
0	5	2	6,673371	2
0	6	2	8,375212	2

The other $f(r,s)$ and $P(r,s)$ values for all s and r combinations are calculated as in third period.

Table 5.9 $f(r,s)$ and $P(r,s)$ values in the 2nd period

s	r	P(r+1,s-1)	P (r,s)	f (r,s)
1	0	1	0	0
	1	1	0	1,8168
	2	2	1	3,5972
	3	2	1	4,665868
	4	2	1	6,673371
	5	2	1	8,375212
2	0	0	0	0,6336
	1	0	0	1,061704
	2	1	0	2,130371
	3	1	0	4,137874
	4	1	0	5,839715
3	0	0	0	0,723741
	1	0	0	1,234487
	2	0	0	3,427816
	3	0	0	6,890709
4	0	0	0	0,995137
	1	0	0	2,042189
	2	0	0	6,145114
5	0	0	0	1,628966
	1	0	0	4,572782
6	0	0	0	2,974566

Then the f -set table for the 2nd period takes the following form:

Table 5.10 f table set for the 2nd period

r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	0,63	0	0,72	0	1	0	1,63	0	2,97	0
1	0	1	1,82	0	1,06	0	1,23	0	2,04	0	4,57	0		
2	1,82	1	3,6	1	2,13	0	3,43	0	6,15	0				
3	3,6	2	4,67	1	4,14	0	6,89	0						
4	4,67	2	6,67	1	5,84	0								
5	6,67	2	8,38	1										
6	8,38	2												

- *Time Period 1*

The necessary calculations for the first period are simpler than all above because there are only one values for s and r . Since this is the first period, there is no patient remaining in the system from previous periods so there is only one s value, $s=0$. Also the number of remaining patients unscheduled is six because no patient has not assigned yet.

Table 5.11 $f(r,s)$ and $P(r,s)$ values in the 1st period

S	R	D	f (r,s)	P (r,s)
0	6	0	11,37521	2
		1	8,361471	
		2	5,657541	
		3	10,53568	
		4	16,92765	
		5	24,84423	
		6	35,79	

The optimum decisions for all time periods, $P(r,s)$, and corresponding $f(r,s)$ values are as in Table 5.12:

Table 5.12 f table set for all time periods

Time Period 4														
r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	1,88	0	5,73	0	11,5	0	18,9	0	26,4	0
1	0	1	4,41	1	10,8	1	19,2	1	29	1	39,1	1		
2	4,41	2	13,3	2	24,2	2	36,6	2	49,3	2				
3	13,3	3	26,8	3	41,7	3	56,9	3						
4	26,8	4	44,2	4	61,9	4								
5	44,2	5	64,5	5										
6	64,5	6												
Time Period 3														
r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	0,45	0	1,08	0	2,51	0	5,43	0	10,1	0
1	0	1	1,82	0	0,61	0	4,42	0	9,28	0	16,4	0		
2	1,82	1	4,24	1	5,05	0	11,2	0	20,3	0				
3	4,24	2	9,83	1	10,7	0	21,3	0						
4	7,59	2	13,2	1	18,5	0								
5	10,7	2	21	1										
6	21	2												
Time Period 2														
r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	0,63	0	0,75	0	1,03	0	1,65	0	2,92	0
1	0	1	1,82	0	1,1	0	1,26	0	2,3	0	4,36	0		
2	1,82	1	3,63	0	2,2	0	3,4	0	5,74	0				
3	3,63	1	4,74	1	4,27	0	6,79	0						
4	4,74	2	6,81	1	6,02	0								
5	6,81	2	8,55	1										
6	8,55	2												
Time Period 1														
r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
6	5,70	2	-	-	-	-	-	-	-	-	-	-	-	-

It is known that in dynamic programming, the decision of a stage cannot be taken unless the previous stage is finished. For example, two patients are assigned to the first period and the number of remaining patients unscheduled is four ($r=4$). If both of the patients are treated in the first period, then the number of patients remaining in the system from previous periods for the second period will be zero ($s=0$). By

looking at the f table set for the second period the optimum number of patients for state (4,0) is two patients. If only one patient can be treated in the first period, then the s value will be one and the optimum decision variable will be one that only one patient will be scheduled to the second period. The decisions for the other periods can be taken in the same way.

5.2.2.2 Alternative Model 2

The second model has three time periods to schedule patients. The model parameters are as defined in the first model.

- *Time Period 3*

The last, 3rd period of this model corresponds to the 4th period of the previous model. All calculations are the same as in that period so f table set for the last period of this model is as follows:

Table 5.13 f table set for the last period

r/s	0		1		2		3		4		5		6	
	$f(r,s)$	$P(r,s)$	$f(r,s)$	$P(r,s)$	$f(r,s)$	$P(r,s)$	$f(r,s)$	$P(r,s)$	$f(r,s)$	$P(r,s)$	$f(r,s)$	$P(r,s)$	$f(r,s)$	$P(r,s)$
0	0	0	0	0	1,88	0	5,73	0	11,5	0	18,9	0	26,4	0
1	0	1	4,41	1	10,8	1	19,2	1	29	1	39,1	1		
2	4,41	2	13,3	2	24,2	2	36,6	2	49,3	2				
3	13,3	3	26,8	3	41,7	3	56,9	3						
4	26,8	4	44,2	4	61,9	4								
5	44,2	5	64,5	5										
6	64,5	6												

- *Time Period 2*

Now, this period corresponds to the third period of the first model and values of $f(r,s)$ and $P(r,s)$ are the same again.

Table 5.14 f table set for the 2nd period

r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	0,45	0	1,08	0	2,51	0	5,43	0	10,1	0
1	0	1	1,82	0	0,61	0	4,42	0	9,28	0	16,4	0		
2	1,82	1	4,24	1	5,05	0	11,2	0	20,3	0				
3	4,24	2	9,83	1	10,7	0	21,3	0						
4	7,59	2	13,2	1	18,5	0								
5	10,7	2	21	1										
6	21	2												

- Time Period 1

Since, there are three time periods the first period of this model is calculated as the first period of the previous model. Again $s=0$ and $r=6$:

Table 5.15 $f(r,s)$ and $P(r,s)$ values in the 1st period

S	R	D	f(r,s)	P(r,s)
0	6	0	24	2
		1	12,5	
		2	8,38	
		3	14,5	
		4	21,4	
		5	29,9	
		6	41	

The optimum decisions for all time periods, $P(r,s)$, and corresponding $f(r,s)$ values are as in Table 4.16. The decisions are taken as described in the first alternative model. Two patients are assigned to the first period. Then, one or zero patients can remain in the system for the second period.

If one patient remains in the system, then one patient can be assigned to the 2nd period, given that there are four patients unscheduled at the beginning of the 2nd period. Because the third period is the last period, the rest of the patients must be scheduled to the 3rd period regardless of how many patients remain in the system.

On the other hand, if none of the patients remain in the system; two patients can be scheduled to the 2nd period according to the f table set. Then two patients must be scheduled to the 3rd period.

Table 5.16 f table set for all time periods

Time Period 3														
r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	1,88	0	5,73	0	11,5	0	18,9	0	26,4	0
1	0	1	4,41	1	10,8	1	19,2	1	29	1	39,1	1		
2	4,41	2	13,3	2	24,2	2	36,6	2	49,3	2				
3	13,3	3	26,8	3	41,7	3	56,9	3						
4	26,8	4	44,2	4	61,9	4								
5	44,2	5	64,5	5										
6	64,5	6												
Time Period 2														
r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
0	0	0	0	0	0,45	0	1,08	0	2,51	0	5,43	0	10,1	0
1	0	1	1,82	0	0,61	0	4,42	0	9,28	0	16,4	0		
2	1,82	1	4,24	1	5,05	0	11,2	0	20,3	0				
3	4,24	2	9,83	1	10,7	0	21,3	0						
4	7,59	2	13,2	1	18,5	0								
5	10,7	2	21	1										
6	21	2												
Time Period 1														
r/s	0		1		2		3		4		5		6	
	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)	f(r,s)	P(r,s)
6	8,55	2												

In practice dynamic programming is not an efficient approach for appointment scheduling because in the dynamic state, it is not possible to take decisions before a stage is finished but in static state it is possible to take all decisions before the clinic session starts. Then, the calculations for dynamic programming approach are used to schedule appointments in static state in the next section.

5.2.3 Static State

A branch-and-bound algorithm developed by Liao, Pegden and Rosenshine (1993) is used and modified to convert dynamic state problem to the static state by using f values calculated for the dynamic programming.

The formulas (11), (12) and (13) given in Chapter three are used for calculating the lower bounds of the branch-and-bound approach. As in the dynamic version idle time of the doctor is ignored while calculating the $h_{wc}(\underline{d}_{i-1})$. The expected idle time of the doctor at stage i is related to three parameters:

j = the number of patients remaining from period $i-2$;

d = number of patients assigned to period $i-1$;

s = number of patients remaining at the beginning of period i ;

The number of patients treated in period i is equal to $(j+d-s)$. Then, the expected idle time of the doctor is:

$$\sum_{s=0}^{D_{i-2}} v_i(s) \sum_{j=1}^{D_{i-2}} (v_{i-1}(j)(1 - (j + d - s)\mu))$$

where $v_{i-1}(j)$ indicates the probability of j and $v_i(s)$ indicates the probability of s .

Then the $h_{tc}(\underline{d}_{i-1})$ takes the form of:

$$h_{tc}(\underline{d}_{i-1}) = c_w \sum_{s=0}^{D_{i-1}} v_{i-1}(s) \sum_{j=1}^{d_{i-1}} \frac{s + j - 1}{\mu} + c_I \sum_{s=0}^{D_{i-2}} v_i(s) \sum_{j=1}^{D_{i-2}} (v_{i-1}(j)(1 - (j + d - s)\mu)) + h_{wc}(\underline{d}_{1-2})$$

The notation is also changed because $h_{wc}(\underline{d}_{i-1})$ refers to only waiting cost of previous periods. By adding the idle time cost now, the total cost of previous periods is shown by $h_{tc}(\underline{d}_{i-1})$. The first term of the formula refers to the waiting cost of customers and the second term refers to the idle time cost of the doctor during period $i-1$ while the last term is the total expected cost from customers scheduled in stages 1 through $i-2$.

5.2.3.1 Static Version of Alternative 1

As it is mentioned in the dynamic state problem, there are four time periods for alternative 1. The c_w , c_i and μ values are as defined in the dynamic state. The g_i values used to calculate the lower bounds are as in Table 5.17.

Table 5.17 g table set for all time periods

Time Period 4								
r/s	0	1	2	3	4	5	6	
0	0	0	1,8793	5,7272	11,5440	18,8820	26,4437	
1	0	4,4147	10,7982	19,1505	29,0240	39,1212		
2	4,4147	13,3337	24,2215	36,6305	49,2632			
3	13,3337	26,7570	41,7015	56,8697				
4	26,7570	44,2370	61,9407					
5	44,2370	64,4762						
6	64,4762							
Time Period 3								
r/s	d3	0	1	2	3	4	5	6
0	0	0	0	0,4489	1,0771	2,5142	5,4336	10,1239
1	0	3	1,8168	0,6128	4,4233	9,2816	16,4179	
	1	0	3,1483	6,1481	10,1207	15,5756	22,8013	
2	0	7,4147	6,2315	5,0497	11,1907	20,2658		
	1	1,8168	4,2390	9,4943	16,8881	26,5598		
	2	3,1483	8,6836	15,1917	23,1821	32,9433		
3	0	16,3337	15,1505	10,6745	21,2920			
	1	6,2315	9,8271	16,2617	27,8723			
	2	4,2390	12,0298	21,9591	34,1663			
	3	8,6836	17,7272	28,2531	44,3531			
4	0	29,7570	28,5738	18,4673				
	1	15,1505	13,2100	26,3630				
	2	7,5852	18,7972	32,9433				
	3	12,0298	24,4946	39,2373				
	4	17,7272	30,7886	45,6208				
5	0	47,2370	46,0538					
	1	28,5738	21,0028					
	2	10,6745	28,8985					
	3	18,7972	35,4788					
	4	24,4946	41,7728					
	5	30,7886	48,1563					
6	0	67,4762						
	1	46,0538						
	2	21,0028						
	3	28,8985						
	4	35,4788						
	5	41,7728						
	6	48,1563						

Time Period 2								
r/s	d2	0	1	2	3	4	5	6
0	0	0	0	0,6336	0,7237	0,9951	1,6290	2,9746
1	0	0	1,8168	1,0617	1,2345	2,0422	4,5728	
	1	0	3,1483	5,7947	8,3093	11,7710	15,6521	
2	0	4,8168	3,6336	2,1304	3,4278	6,1451		
	1	1,8168	3,5972	6,3055	9,9410	14,7148		
	2	3,1483	8,3302	13,6726	19,3774	25,7940		
3	0	7,2390	6,0558	4,1379	6,8907			
	1	3,6336	4,6659	8,4988	13,7516			
	2	3,5972	8,8410	15,0120	22,3213			
	3	8,3302	16,2081	24,4484	33,4005			
4	0	10,5852	9,4020	5,8397				
	1	6,0558	6,6734	11,9617				
	2	4,6659	11,0343	18,8226				
	3	8,8410	17,5475	27,3923				
	4	16,2081	26,9839	38,4715				
5	0	13,6745	12,4913					
	1	9,4020	8,3752					
	2	6,6734	14,4972					
	3	11,0343	21,3581					
	4	17,5475	29,9278					
	5	26,9839	41,0070					
6	0	24,0028						
	1	12,4913						
	2	8,3752						
	3	14,4972						
	4	21,3581						
	5	29,9278						
	6	41,0070						
Time Period 1								
r/s	d1	0	1	2	3	4	5	6
6	0	11,3752						
	1	8,3615						
	2	5,6575						
	3	10,5357						
	4	16,9277						
	5	24,8442						
	6	35,7900						

The stages start from $i=0$ and for $i=0$,

$d_0=0$;

$h=0$;

Then i is increased by one and for $i=1$:

$v_1(0) = 1$ since there are no customers remaining in the system from previous periods. Also, because it is the first period, there is no cost of schedules of previous periods and lower bound is related to only the expected cost of decisions from period 1 through 4. Then, the lower bound values of the first period are equal to the values in the g table and the first branch is as follows:

Decision	d_1
	<u>[0]</u>
	11,552
	<u>[1]</u>
	8,498457
	<u>[2]</u>
	5,709824
	<u>[3]</u>
	10,59739
	<u>[4]</u>
	16,94907
	<u>[5]</u>
	24,92665
	<u>[6]</u>
	35,81178

where values between [...] indicates the possible values of decision d_1 and the values under the line refers to the lower bound of each branch.

Now, it is necessary to decide from which branch we should continue. Since, lower bounds can decrease in the second period; next calculations are made for each branch.

To make the decision for the second stage ($i=2$) the calculations made are as follows:

For $d_1 = 0, 1, 2, \dots, 6$ possible d_2 values are evaluated according to the corresponding lower bound.

➤ $d_1=0$:

$$v_2(s) = \sum_{j=(s-0)^+}^0 v_1(j)P(s|0+j) \quad s=0$$

$$v_2(0) = 0;$$

$$h_{tc}(0) = c_w \sum_{s=0}^0 v_1(s) \sum_{j=0}^0 \frac{s+j-1}{\mu} + c_I(1-d_i\mu) + 0$$

$$h_{tc}(0) = 3;$$

Lower bounds are calculated by all possible d_2 values:

- $d_2=0$;

$$LB(0,0) = \sum_{s=0}^0 v_2(s)g_2(6,s,0) + h_{tc}(0)$$

$$LB(0,0)=27;$$

- $d_2=1$;

$$LB(0,1) = \sum_{s=0}^0 v_2(s)g_2(6,s,1) + h_{tc}(0)$$

$$LB(0,1)=15,49127;$$

- $d_2=2$;

$$LB(0,2) = \sum_{s=0}^0 v_2(s)g_2(6,s,2) + h_{tc}(0)$$

$$LB(0,2)=11,37521;$$

- $d_2=3$;

$$LB(0,3) = \sum_{s=0}^0 v_2(s) g_2(6, s, 3) + h_{tc}(0)$$

$$LB(0,3) = 17,4972;$$

- $d_2=4$;

$$LB(0,4) = \sum_{s=0}^0 v_2(s) g_2(6, s, 4) + h_{tc}(0)$$

$$LB(0,4) = 24,3581;$$

- $d_2=5$;

$$LB(0,5) = \sum_{s=0}^0 v_2(s) g_2(6, s, 5) + h_{tc}(0)$$

$$LB(0,5) = 32,92775;$$

- $d_2=6$;

$$LB(0,6) = \sum_{s=0}^0 v_2(s) g_2(6, s, 6) + h_{tc}(0)$$

$$LB(0,6) = 44,00702;$$

The calculations for other d_1 values are made in the same way and the results for the second period are as in Table 5.18:

Table 5.18 Lower bound values for the second period of alternative 1

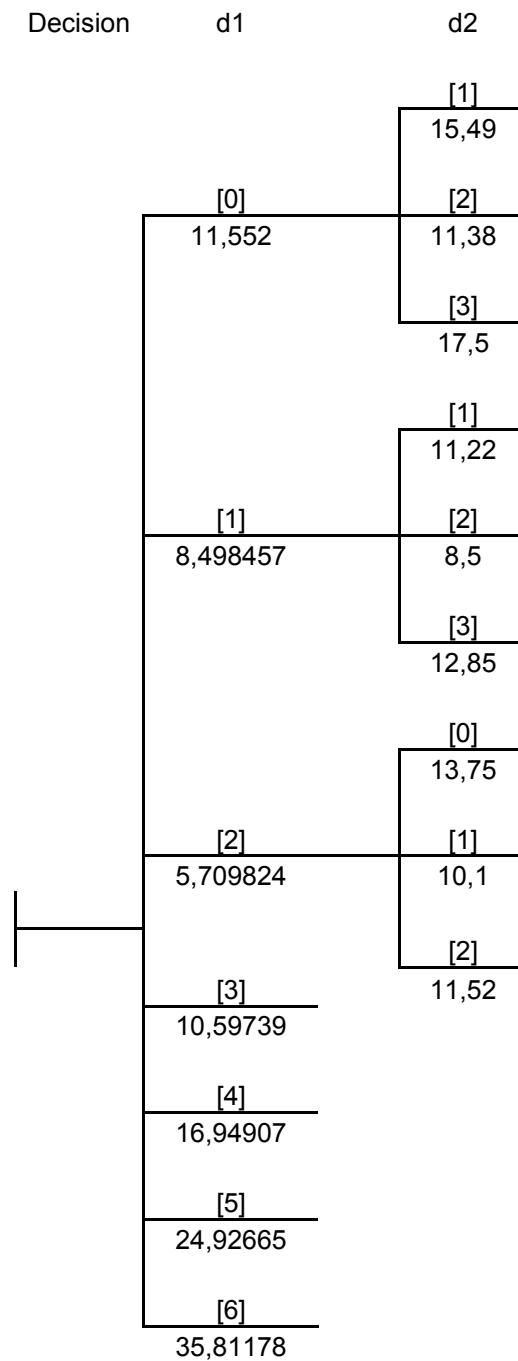
d1	d2	LB(d1,d2)
0	0	27,00280474
	1	15,49127104
	2	11,37521153
	3	17,49719938
	4	24,35809571
	5	32,92775133
	6	44,00702081
1	0	15,49127104
	1	11,21880049
	2	8,490170738
	3	12,85110728
	4	19,36429062
	5	28,80073541
2	0	13,75429745
	1	10,10436686
	2	11,52297505
	3	16,8399345
	4	25,21766558
3	0	14,24456764
	1	13,96272659
	2	17,46777994
	3	24,61192035
4	0	19,2711417
	1	21,42758643
	2	27,36667637
5	0	27,44230397
	1	33,20222178
6	0	39,61935782

For the next period, it is unnecessary to make all calculations for each branch. As it is seen on the table, the lower bounds for $d_1 = 3, 4, 5$ and 6 are too high that it will be unnecessary to continue calculations for these branches.

For $d_1 = 2$, the calculations of the next period can be made for $d_2 = 0, 1$ and 2 because they have smaller lower bounds.

For $d_1 = 0$ and 1, lower bounds of $d_2 = 1, 2$ and 3 are smaller than others so these branches are continued to calculate.

Then, the new branch-and-bound scheme is as follows:



Now, the calculations are made for the third period as in the second period.

$$\triangleright d_1=0 \text{ \& } d_2=1;$$

$$v_3(s) = \sum_{j=(s-1)^+}^0 v_2(j) P(s|1+j) \quad s = 0,1$$

$$v_2(0) = 1$$

$$v_2(1)=0;$$

$$h_{tc}(0,1) = c_w \sum_{s=0}^0 v_2(s) \sum_{j=1}^1 \frac{s+j-1}{\mu} + c_I (1 - (d_1 + d_2)\mu) + h_{Tc}(0)$$

$$h_{tc}(0,1) = 4,8168;$$

Lower bounds are calculated by all possible d_3 values:

- $d_3=0;$

$$LB(0,1,0) = \sum_{s=0}^1 v_3(s) g_3(5, s, 0) + h_{tc}(0,1)$$

$$LB(0,1,0)=52,054;$$

- $d_3=1;$

$$LB(0,1,1) = \sum_{s=0}^1 v_3(s) g_3(5, s, 1) + h_{tc}(0,1)$$

$$LB(0,1,1)=33,391;$$

- $d_3=2;$

$$LB(0,1,2) = \sum_{s=0}^1 v_3(s) g_3(5, s, 2) + h_{tc}(0,1)$$

$$LB(0,1,2)=15,4913;$$

- $d_3=3$;

$$LB(0,1,3) = \sum_{s=0}^1 v_3(s) g_3(5, s, 3) + h_{tc}(0,1)$$

$$LB(0,1,3) = 23,614;$$

- $d_3=4$;

$$LB(0,1,4) = \sum_{s=0}^1 v_3(s) g_3(5, s, 4) + h_{tc}(0,1)$$

$$LB(0,1,4) = 29,3114;$$

- $d_3=5$;

$$LB(0,1,5) = \sum_{s=0}^1 v_3(s) g_3(5, s, 5) + h_{tc}(0,1)$$

$$LB(0,5) = 35,6054$$

The calculations for other d_1 and d_2 values are made in the same way and the results for the third period are as in Table 5.19:

Table 5.19: Lower bound values for the third period of alternative 1

d1	d2	d3	LB(d1,d2,d3)
0	1	0	52,05376456
		1	33,39055024
		2	15,49127104
		3	23,61398558
		4	29,31139639
		5	35,60539472
	2	0	35,9260472
		1	20,94975303
		2	19,80775994
		3	24,86418602
		4	30,85295352
	3	0	25,51279331
		1	22,09690349
		2	24,0087557
		3	29,47487412

1	1	0	33,39055024
		1	18,78413106
		2	11,21880049
		3	15,66337751
		4	21,36078832
	2	0	21,31962802
		1	13,55123415
		2	13,60756086
		3	18,66398694
1	3	0	16,47489503
		1	15,34176029
		2	19,13745705
2	0	0	35,9260472
		1	21,31962802
		2	13,75429745
		3	18,19887447
		4	23,89628528
	1	0	22,46162173
		1	13,11941377
		2	11,79409306
		3	16,43791267
	2	0	16,04670819
		1	13,80669469
		2	17,21603609

The branches that have smaller lower bounds are:

$$d_1 = 0, d_2 = 1, d_3 = 2$$

$$d_1 = 1, d_2 = 1, d_3 = 1 \text{ \& } d_3 = 2 \text{ \& } d_3 = 3$$

$$d_1 = 1, d_2 = 2, d_3 = 1 \text{ \& } d_3 = 2$$

$$d_1 = 2, d_2 = 0, d_3 = 2 \text{ \& } d_3 = 3$$

$$d_1 = 2, d_2 = 1, d_3 = 1 \text{ \& } d_3 = 2$$

$$d_1 = 2 \text{ \& } d_2 = 2 \text{ \& } d_3 = 1$$

Then, the new branch-and-bound scheme is as follows:

Decision	d1	d2	d3
		[1] 15,49	[2] 15,49127
	[0] 11,552	[2] 11,38	
		[3] 17,5	[1] 18,78413
		[1] 11,22	[2] 11,2188
			[3] 15,66338
			[1] 13,55123
	[1] 8,498457	[2] 8,5	[2] 13,60756
		[3] 12,85	[2] 13,7543
		[0] 13,75	[3] 18,19887
			[1] 13,11941
	[2] 5,709824	[1] 10,1	[2] 11,79409
		[2] 11,52	[1] 13,80669
	[3] 10,59739		
	[4] 16,94907		
	[5] 24,92665		
	[6] 35,81178		

For the last, fourth, period, all of the remaining patients must be scheduled so there is only one available d_4 value for each branch.

The calculations for the 4th period are the same as in the previous periods and the results are as follows:

Table 5.20: Lower bound values for the last period of alternative 1

d1	d2	d3	d4	LB(d1,d2,d3,d4)
0	1	2	3	28,45299553
1	1	1	3	18,78413106
		2	2	16,15106906
		3	1	17,02438878
	2	1	2	13,96602428
		2	1	15,19887275
2	0	2	2	13,75429745
		3	1	19,55988574
	1	1	2	13,15100961
		2	1	13,45644675
	2	1	1	14,19234402

Then, the branch-and-bound scheme for the first alternative is given below. As it is seen on the scheme, the minimum lower bound for the last period is obtained by the sequence of decisions (2, 1, 1, 2). The lower bound of this sequence is 13,15.

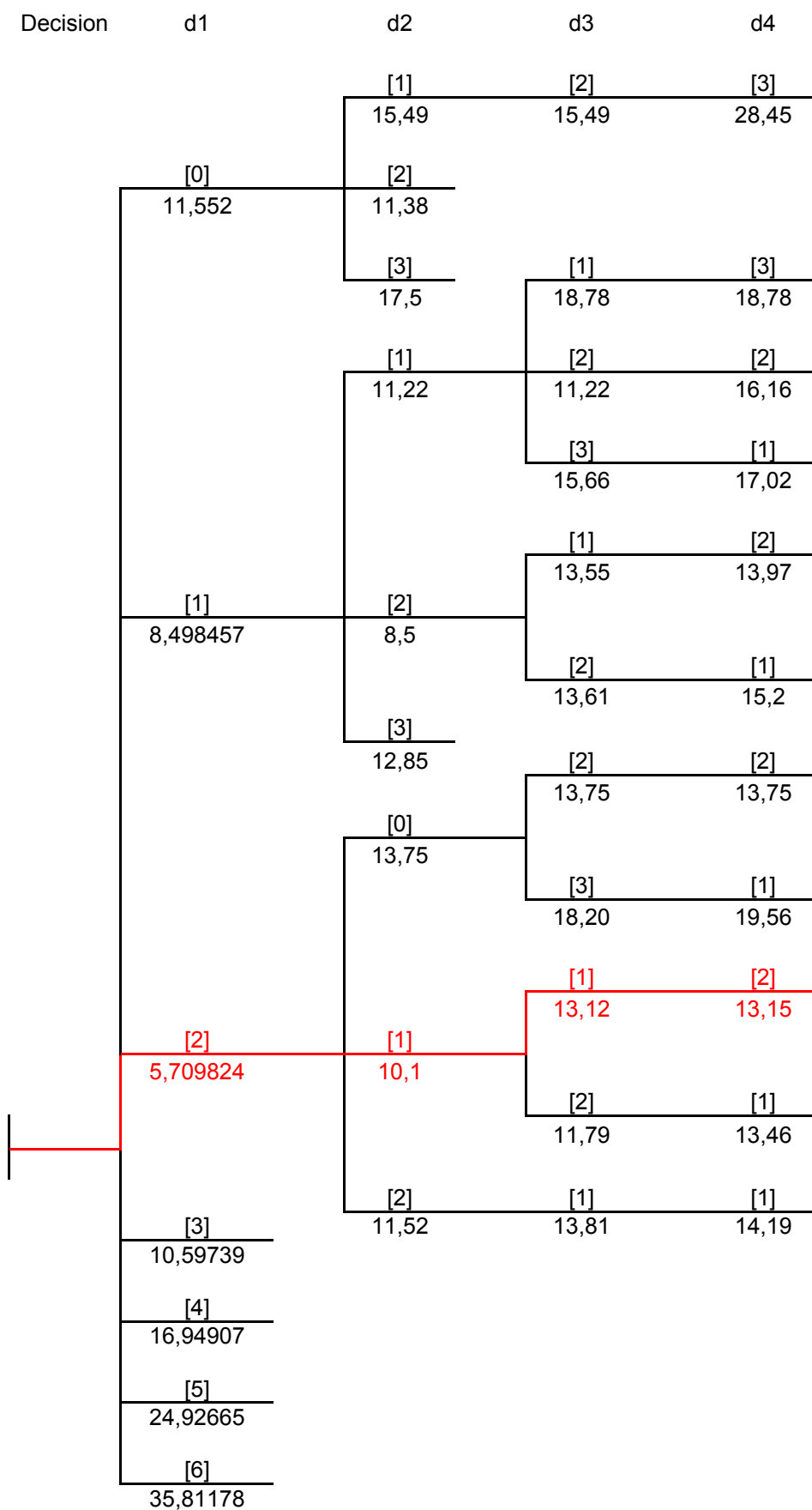
As a result, for the first alternative model optimum number of patients to be scheduled at the beginning of each period is as follows:

1st period: 2 patients

2nd period: 1 patient

3rd period: 1 patient

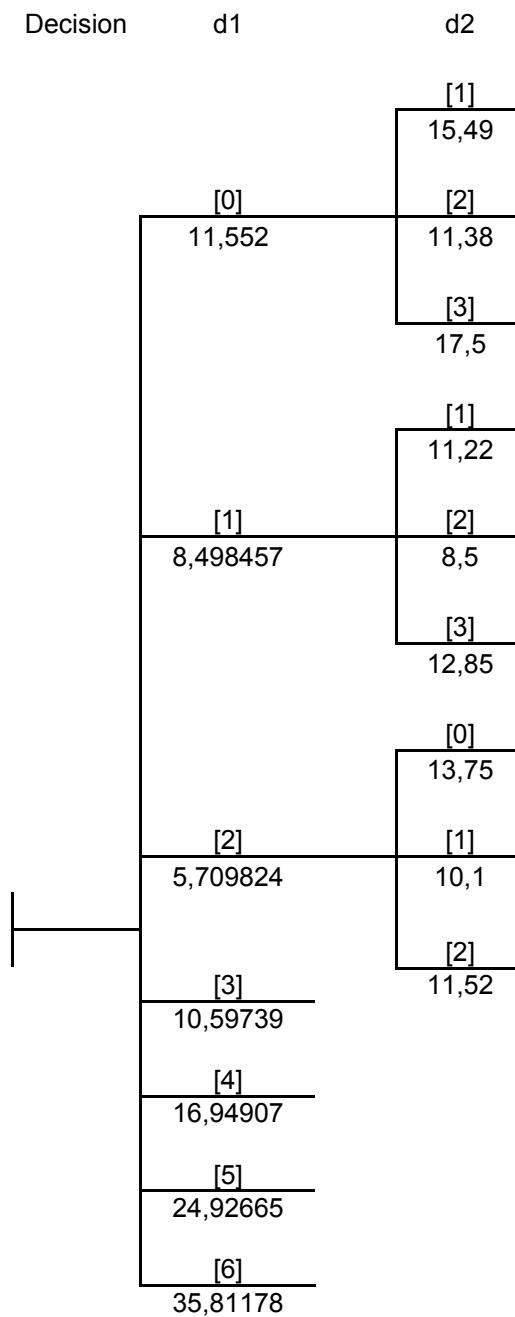
4th period: 2 patients

Figure 5.1 Branch-and-bound scheme for the 1st alternative

5.2.3.2 Static Version of Alternative 2

As it is mentioned in the previous section, there are three time periods for alternative model 2. The c_w , c_i and μ values are still as defined in the dynamic state. The g_i values in Table 4.17 are also used for the calculation of lower bounds.

Lower bounds for the first and second period are the same as in the alternative model 1. Then, the branch-and-bound scheme is:



The calculations for the 3rd period are the same as in the previous model and the results are as follows:

Table 5.21: Lower bound values for the last period of alternative 2

d1	d2	d3	LB(d1,d2,d3)
0	1	5	35,60539472
	2	4	30,85295352
	3	3	29,47487412
1	1	4	21,36078832
	2	3	18,66398694
	3	2	19,13745705
2	0	4	23,89628528
	1	3	16,43791267
	2	2	17,21603609

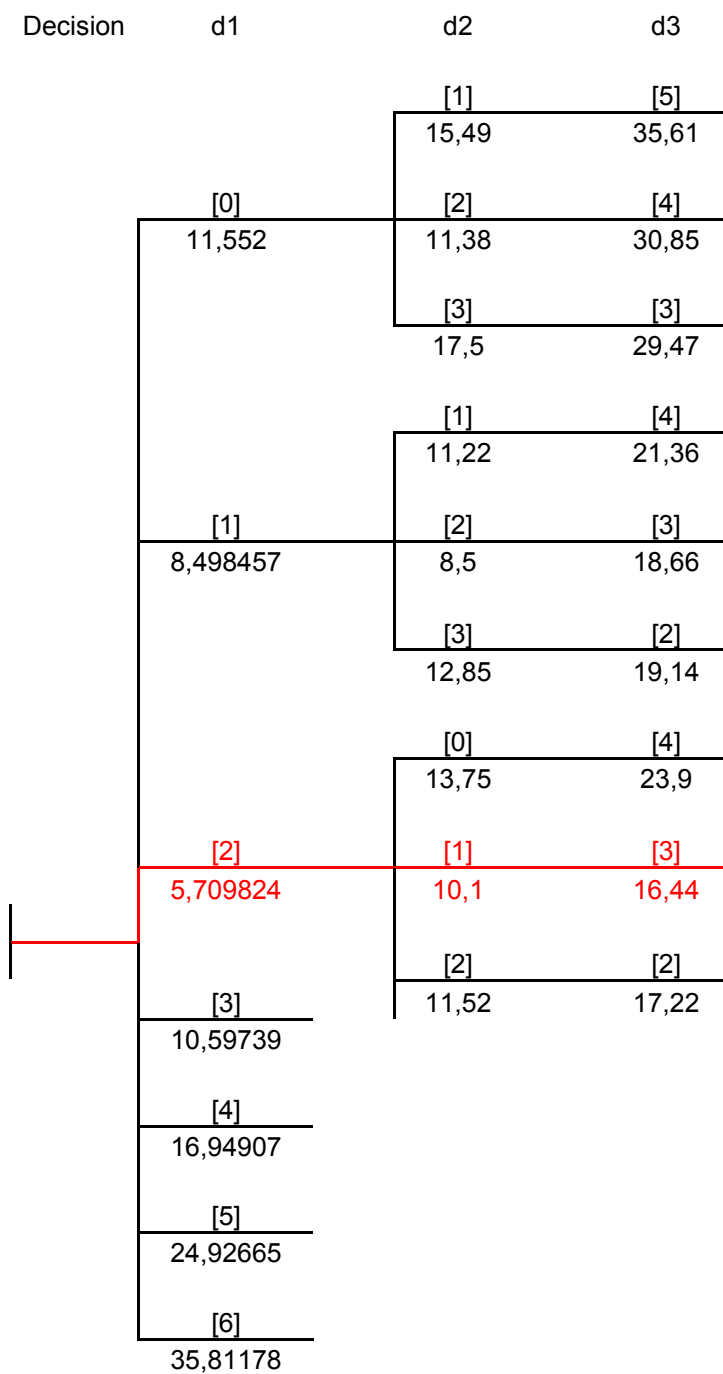
Then, the branch-and-bound scheme for the second alternative is given below. As it is seen on the scheme, the minimum lower bound for the last period is obtained by the sequence of decisions (2 , 1, 3). The lower bound of this sequence is 16,438.

As a result, for the second alternative model optimum number of patients to be scheduled at the beginning of each period is as follows:

1st period: 2 patients

2nd period: 3 patients

3rd period: 1 patient

Figure 5.2 Branch-and-bound scheme for the 2nd alternative

Finally, the results for appointment scheduling problem in DEU, Dept.of Ort. and Trau. is as follows:

	Alternative Model 1		Alternative Model 2	
	Dynamic Version	Static Version	Dynamic Version	Static Version
1. Period	2	2	2	2
2. Period	1 or 2	1	1 or 2	1
3. Period	1 or 2	1	2 or 3	3
4. Period	1 or 2	2	-	-

For the dynamic state of both alternative model 1 and alternative model 2, two patients are assigned to the first period. The assignments of the latter periods cannot be known exactly before the previous period finishes because of the characteristic of dynamic programming approach.

As it is mentioned above, dynamic state is not used commonly in practice because it is necessary to determine the decisions of all stages before the clinic session starts. Therefore, static state solution is more efficient for DEU, Dept. of Ort and Trau.

For the static state with four time periods optimum patient assignment at the beginning of each period is (2, 1, 1, 2). If the clinic session is divided into three time periods, then optimum assignment takes the form of (2, 1, 3) patients.

In fact, it is the responsibility of the management to determine the number of periods but alternative model 2 with three time periods seems more efficient for the doctor. It enables him/her to save more time.

5.3 Personnel Scheduling Problem

As it is mentioned in the previous section, a part of the team treats patients on an appointment based system and the rest of the team treats patients in the outpatient clinic. In the outpatient clinic, there is not a rule for sequencing patients. Patients arrive randomly and all patients enter the clinic within a day are consulted.

The patient flow in the outpatient clinic is not so complex that when a patient enters the system, after being registered by a registrar, he / she waits for consultation and then exits. The system is a multi-server system that the number of servers may change during the time of the day according to the patient arrival rates.

In the current system, there are 24 residents working in the outpatient clinic but there is not a system to schedule the working hours of these doctors. A doctor works in the clinic only one day a week but the working hours of the doctor within a day are not specified. The doctor must be in the clinic throughout the whole day.

In this study, scheduling systems is proposed with limited staff size. Also, different work patterns and work hours are analyzed in order to obtain the best schedule for doctors.

A clinic session is divided into eight time periods that each one last for an hour. The clinic works for eight hours per day and five days a week. But, it is not necessary that all staff must work eight hours a day and five days a week. The objective of the study is to determine optimal work hours of doctors.

As it is described in Chapter four, formulation of Ozkarahan (1991) are used to form the linear program but the right-hand sides of the formulation are not known. To determine the doctor requirements per time period, right-hand sides, queuing theory is used. Therefore, in the next section use of queuing theory for personnel scheduling is given.

5.3.1 Queuing Theory

In this study, queuing theory is used to determine minimum number of doctors per time period in order to meet the patient demand. As it is mentioned in the previous chapter, the performance measure of the system is patient waiting time and the management wants to keep it within limits.

As the first step, two parameters of the queuing theory, λ and μ , are needed to be calculated. According to the data and interviews with personnel, input analysis is performed to determine patient arrival rate, λ , and doctor service rate, μ .

5.3.2 Input Analysis

The first step of the study is analyzing the data. The type of the data needed was patient arrivals to the system and the service times of the servers. The arrival of each patient to the system was recorded but there was not any record about consultation times.

5.3.2.1 Steps of Input Analysis

There are four steps in the development of a useful analysis of input data:

➤ *Collecting Data*

Collect data from the real system of the interest. This often requires a substantial time and resource commitment. Unfortunately, in some situations it is not possible to collect data. When data are not available, expert opinion and knowledge of the process must be used to make educated guesses.

➤ *Using Data*

Identify a probability distribution to represent the input data. When data are available, this step typically begins by developing a frequency distribution or histogram of the data. Based on the frequency distribution or histogram, a family of distributions is chosen. Several well-known distributions often provide good approximations in practice.

➤ *Parameter Estimation*

Choose parameters that determine a specific instance of the distribution family. When data are available, these parameters may be estimated from the data.

➤ *Goodness-of-fit*

Evaluate the chosen distribution and the associated parameters for goodness-of-fit. Goodness-of-fit may be evaluated informally via graphical methods or formally via statistical tests. The chi-square and the Kolmogorov-Smirnov tests are standard goodness-of-fit tests. If not satisfied that the chosen distribution is a good approximation of the data, then second step is repeated and a different family of distribution is chosen.

5.3.2.2 Patient Arrivals

➤ *Data Collection*

Data about patient arrivals is collected from the Orthopedic Department of Dokuz Eylul University for 23 weeks. Arena 5.0 (2000) is used as a tool to analyze the input data and it was necessary minimum 20 data points to fit data to a distribution. The data could not be recorded on some days because of holidays, strike etc. so 23 weeks are recorded for input analysis. Arrival of each patient to the system is recorded. Some examples of these records are shown in Appendix-A. The second column,

Date, refers to the day on which arrivals recorded. *Hour* indicates the time patient enters the system. *Patient File* is the document number of the patient. This document number is identical for each patient and used among all departments. The documents of patients are stored in the archives and they are requested when the patient is registered. *User* is the registrar who makes the registration of the patient to the system. Sometimes, the patient may be directed from other departments for a necessary consultation. *Institution* refers to the social insurance of the patient. The last column, *Status* indicates whether the patient is a new patient or not. If it is a new patient a new document is formed for him / her so it needs more time to be registered.

The total number of patient arrivals during each day and time period is shown in Appendix-B. The last column indicates the total arrivals during the day. As it is seen by Figure 5.3 more patients arrive in the system between 8 a.m. and 9 a.m. During the period 16 p.m.-17 p.m. least people apply for consultation. By looking at the Figure 5.4 it is obvious that the busiest day of the week is Monday. The application for consultation decreases throughout the week and it is the least on Friday so it is necessary to analyze each day of the week separately.

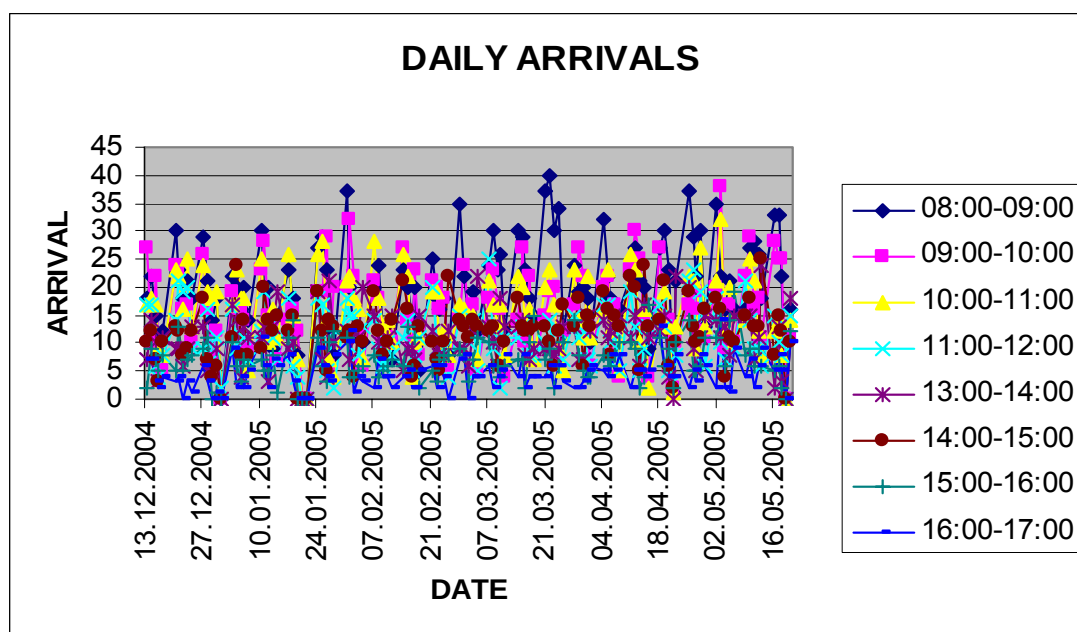


Figure 5.3 Daily patient arrivals per hour

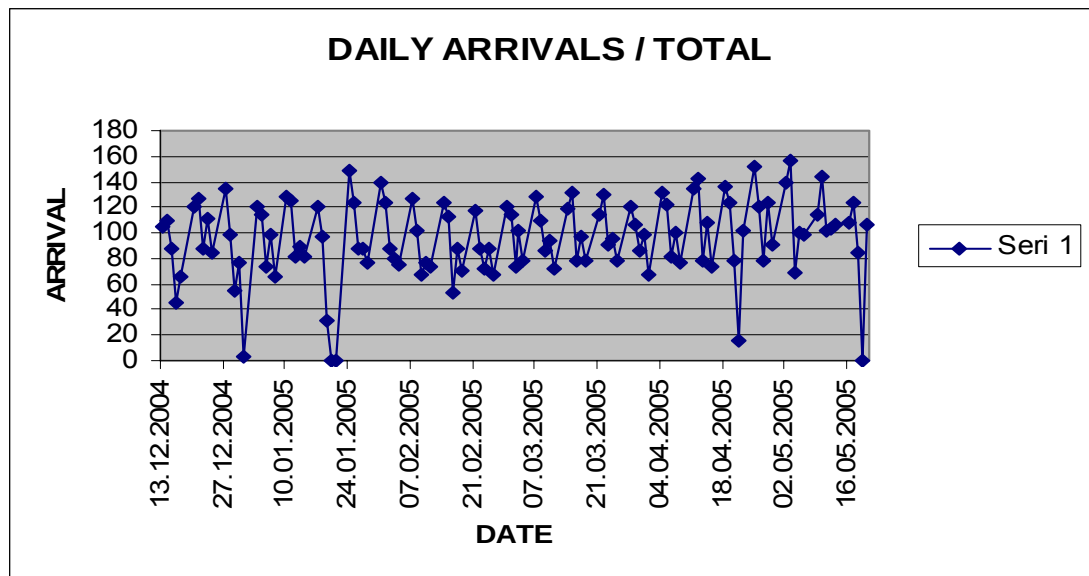


Figure 5.4 Total daily Arrivals

As it is seen there are some great decreases on some days. These data points were existed because of some extraordinary situations. For some days there was not any record because of the public holidays. The first exception is recorded on 31.12.2004 Because it was the Christmas only 3 patients applied for consultation in the morning but also no patients applied in the afternoon because of public holiday. On 19.01.2005 afternoon there was an official holiday so no patients were treated 20.01.2005, 21.01.2005 and 19.05.2005 were also public holidays so there was no patient arrival on these days.

On the other hand, there is a great decrease on some days such as 21.04.2005 although it was not a public holiday. This is because there was a strike among the doctors and other medical personnel so only emergencies are consulted.

The distribution of arrivals according to weeks is as in Appendix C respectively according to the days of week and time of day. The first graphic shows that the busiest period of the day is during the morning. Then, the arrival of the patients decreases, generally. The second graphic indicates the busiest day of the week as Monday. Also, the effects of holidays and the strike on 21.04.2005 are seen on this graphic.

Each time period of the day and each day are analyzed separately and total arrivals according to days and time periods are as in Appendix D. The graphics of hourly arrivals and total arrivals also can be seen on the appendix. The effects of the public holidays and the strike on 21.04.2005 are also obvious on the graphics.

➤ *Using Data*

Histograms: To organize raw data it must be first put in the form of a histogram. A histogram is constructed as follows:

- i. Divide the range of the data into intervals.
- ii. Label the horizontal axis to conform to the intervals selected.
- iii. Determine the frequency of occurrences within each interval.
- iv. Label the vertical axis so that the total occurrences can be plotted for each interval.
- v. Plot the frequencies on the vertical axis (Banks et. al., 2001)

To test the compatibility of a data set, first the theoretical distribution must be identified. Because it is always assumed that arrivals are distributed by Poisson, we try to fit the data set to Poisson distribution. A Poisson distribution is a discrete distribution and is used to describe the number of events that occur in an interval of time when the events occur independently. The software used for input analysis, Arena 5.0, constructs histogram automatically when a distribution is fit.

As it is mentioned by the graphics, patient arrivals differ according to the period of the day and the day of the week. So it is necessary to divide data according to the arrival time of the patient. The arrivals on public holidays and 21.04.2005 are extraordinary points so these data is eliminated.

➤ *Parameter Estimation*

Once the form of a distribution has been hypothesized, the next step is to estimate the parameters of the distribution. In many cases the sample mean or sample mean and variance provide basis for estimating the parameters of the selected distribution. For the Poisson distribution, if the observations in a sample of size n are $x_1, x_2, \dots, x_n$, the sample mean and variance are defined by:

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \text{ (Pegden et. al, 1990)}$$

➤ *Goodness-of-Fit Tests*

Once a distribution is selected and the parameters of the selected distribution are estimated it is needed to assess the quality of the fit. This can be done by using formal statistical tests or by employing a simple graphical method in which an overlay of the theoretical distribution is displayed on a histogram of the data and a visual assessment is made to determine the quality of fit.

To test statistically the hypothesis, one of the goodness-of-fit tests is used. A goodness-of-fit test attempts to measure and evaluate the deviation of the sample distribution from the theoretical.

Chi-square Test

One of the most widely used measures or tests of the discrepancy between an observed and an expected set of data is the X^2 statistic. The Chi-square statistic is calculated as follows:

$$X^2 = \frac{\sum_{e=1}^k (f_o - f_e)^2}{f_e}$$

where

K= number of classes or intervals

F_o = observed frequency for each class or interval

F_e = expected frequency for each class or interval predicted by the theoretical distribution

The hypotheses are:

H₀: the random variable, X, conforms to the distributional assumption with the parameter(s) given by the parameter estimate

H₁: the random variable, X, does not conform

The degrees of freedom, v, used in the test is:

$$v = k - p - 1$$

where

p is the number of population parameters used in the calculation of the theoretical frequencies, which were estimated from the observed sample. The number of parameters estimated for Poisson distribution (p) is 1.

If the calculated value of X^2 is greater than the tabulated value (at a given level of significance and appropriate degrees of freedom), the H₀ is rejected.

There is also another goodness-of-fit test called Kolmogorov-Smirnov test but the K-S test is only valid for testing continuous distributions. Since Poisson is a discrete distribution K-S test cannot be applied to the data set for goodness-of-fit.

5.3.2.3 Fitting Input Distributions Via The Input Analyzer Of Arena 5.0

The Input Analyzer is a standard tool that accompanies Arena and is designed specifically to fit distributions to observed data. It provides numerical estimates of the appropriate parameters or fit a number of distributions to the data and selects the most appropriate one. When the Input Analyzer fits a distribution to the data, it estimates the distribution's parameters and calculates a number of measures of how good the distribution fits the data.

The data file is loaded to the Input Analyzer as a text file and then a histogram of the data set is drawn by the analyzer. The “*Fit*” menu of the analyzer provides options for fitting individual probability distributions to the data. After a distribution is fitted, its density function is drawn on the top of the histogram display and a summary of the characteristics of the fit is displayed in the text section.

The other measure of a distribution's fit to the data is the Chi-square goodness-of-fit tests. The Input Analyzer reports information about the test in the text window. Of particular interest is the p-value which will always fall between 0 and 1. To interpret this value, larger p-values indicate better fits. Corresponding p-values less than about 0.05 indicate that the distribution is not a very good fit. Of course, as with any statistical hypothesis test, a high p-value does not constitute “proof” of a good fit- just a lack of evidence against the fit (Kelton et. al., 2002).

Because Poisson distribution is selected for fitting the data, the Input Analyzer performs only Chi-square test and the results are as shown in Appendix E. The results of goodness-of-fit tests are also can be seen in the appendix. The first three columns indicate the sample mean, sample standard deviation and the square error of the related data set respectively. The fourth and fifth columns refer to the parameters of Chi-square test and they are determined by the Input Analyzer. The last two columns show the results of goodness-of-fit tests.

5.3.2.4 Service Times

Because there was not any record about the service times of patients in the system, it is necessary to make input analysis in order to no existing data. There are a number of possibilities for quantifying the probabilistic estimates given by people depending on what estimated values are available. According to the interviews made with doctors and registrars only the mean value of the service times is known.

If only the mean value can be estimated, then there are several options. First; mean value may itself be used directly as a constant time delay. However, this option should be used only when there is a very small variability in the actual process. On the other hand, it may be justifiable to use an exponential distribution with given estimate as the mean value. If the mean service rate is known (i.e. patients served per hour), then the time between servicing patients can be modeled as exponentially distributed since rate data can be justified as being Poisson distributed (Pegden et. al, 1990).

The interviews with related people show that the average consultation time per patient is approximately 10 minutes. As it is assumed that service times are distributed exponentially, it can be justifiable to use exponential distribution with a mean of 10 minutes for consultation times.

5.3.3 Queuing System

After determining patient arrival rates per hour and service rate of the doctor, queuing formulas can be applied to the data. As a result of the input analysis, patient arrivals fit Poisson distribution and interarrival times of patients are distributed exponentially. The service times are assumed to fit exponential distribution. Then the system notation may be $M / M / s$ where “s” indicates the number of doctors needed. By taking into the consideration patient waiting times, the number of doctors is determined by using queuing formulas. The calculations are made for each day and each time period separately.

First, for the stability of the system, the following condition must be satisfied:

$$\frac{\lambda}{s\mu} \leq 1$$

Then, by using the “s” value, queuing formulas are used to calculate the performance measures of the system. The main performance measure for this system is the patients’ waiting time so it is not acceptable when waiting time of patients exceed 30 minutes. The calculations are as follows:

5.3.3.1 Monday

➤ Time Period-1 (08:00-09:00)

$\lambda = 28$ patients / hour

$\mu = 6$ patients / hour

To satisfy the first constraint $\frac{28}{s * 6} \leq 1$ there must be at least five doctors during the period. Then “s” will be five and the other formulas mentioned in Chapter four are calculated. The results are as in Table 5.22.

Table 5.22 Performance measures of the system between 08:00-09:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,003031097	L_q	0,782673 patients
P_1	0,014145118	W_q	0,027953 = 1,677157 min.
P_2	0,033005276	W_s	0,194619 = 11,67716 min.
P_3	0,051341541	L_s	5,44934 patients
P_4	0,059898464	ρ	0,933333
P_5	0,055905234		

As it is seen on the table, patient waiting time, 1,677157 minutes, is short so that five doctors will be enough on Monday between 08.00 and 09:00 to meet the demand while keeping patient waiting time under desired level.

➤ *Time Period-2 (09:00-10:00)*

$\lambda = 21$ patients / hour

$\mu = 6$ patients / hour

$$\frac{21}{s * 6} \leq 1$$

There must be at least four doctors during the period. The calculations for $s=4$ are as in Table 5.23.

Table 5.23 Performance measures of the system between 09:00-10:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,014751076	L_q	0,645628 patients
P_1	0,051628765	W_q	0,030744= 1,844653 min.
P_2	0,090350338	W_s	0,197411= 11,84465 min.
P_3	0,105408728	L_s	4,145628 patients
P_4	0,092232637	ρ	0,875000
P_5	0,322814229		

Between 09:00 and 10:00 on Monday, patient waiting time, 1,844653 minutes, is short enough that four doctors can meet the demand while keeping patient waiting time under desired level.

➤ *Time Period-3 (10:00-11:00)*

$\lambda = 21$ patients / hour

$\mu = 6$ patients / hour

$$\frac{21}{s * 6} \leq 1$$

There must be at least four doctors during the period. The calculations for $s=4$ are as in Table 5.24.

Table 5.24 Performance measures of the system between 10:00-11:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,014751076	L_q	0,645628 patients
P_1	0,051628765	W_q	0,030744= 1,844653 min.
P_2	0,090350338	W_s	0,197411= 11,84465 min.
P_3	0,105408728	L_s	4,145628 patients
P_4	0,092232637	ρ	0,875000
P_5	0,322814229		

As it is seen, on Monday four doctors are enough between 10:00 and 11:00 to keep patient waiting time between desired levels.

➤ *Time Period-4 (11:00-12:00)*

$$\lambda = 16 \text{ patients / hour}$$

$$\mu = 6 \text{ patients / hour}$$

$$\frac{16}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.25.

Table 5.25 Performance measures of the system between 11:00-12:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,028037383	L_q	0,708896 patients
P_1	0,074766355	W_q	0,044306= 2,658359 min.
P_2	0,099688474	W_s	0,210973= 12,65836 min.
P_3	0,088611976	L_s	3,375562 patients
P_4	0,118149302	ρ	0,888889
P_5	0,315064805		

As it is seen on the table, patient waiting time of 2,658359 minutes is acceptable so that three doctors will be enough on Monday between 11.00 and 12:00 to meet the demand while keeping patient waiting time under desired level.

➤ *Time Period-5 (13:00-14:00)*

$\lambda = 12$ patients / hour

$\mu = 6$ patients / hour

$$\frac{12}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.26.

Table 5.26 Performance measures of the system between 13:00-14:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,111111111	L_q	0,296296 patients
P_1	0,222222222	W_q	0,024691= 1,481481 min.
P_2	0,222222222	W_s	0,191358= 11,48148 min.
P_3	0,148148148	L_s	2,296296 patients
P_4	0,148148148	ρ	0,666666
P_5	0,296296296		

Between 13:00 and 14:00 on Monday, patient waiting time of 1,481481 minutes is short enough that three doctors can meet the demand while keeping patient waiting time acceptable.

➤ *Time Period-6 (14:00-15:00)*

$\lambda = 15$ patients / hour

$\mu = 6$ patients / hour

$$\frac{16}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.27.

Table 5.27 Performance measures of the system between 14:00-15:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,04494382	L_q	0,585206 patients
P_1	0,112359551	W_q	0,039014= 2,340824 min.
P_2	0,140449438	W_s	0,20568= 12,34082 min.
P_3	0,117041199	L_s	3,085206 patients
P_4	0,146301498	ρ	0,833333
P_5	0,365753745		

Since 2,340824 minutes is an acceptable waiting time for patients, three doctors are enough between 14:00 and 15:00 on Monday.

➤ *Time Period-7 (15:00-16:00)*

$\lambda = 9$ patients / hour

$\mu = 6$ patients / hour

$$\frac{9}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.28.

Table 5.28 Performance measures of the system between 15:00-16:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,142857143	L_q	0,482143 patients
P_1	0,214285714	W_q	0,053571= 3,214286min.
P_2	0,160714286	W_s	0,220238= 13,21429min.
P_3	0,241071429	L_s	1,982143 patients
P_4	0,180803571	ρ	0,750000
P_5	0,090401786		

Patient waiting time between 15:00 and 16:00 is a bit longer than the previous time periods but it is still acceptable so two doctors can meet the demand.

➤ *Time Period-8 (16:00-17:00)*

$\lambda = 6$ patients / hour

$\mu = 6$ patients / hour

$$\frac{6}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.29.

Table 5.29 Performance measures of the system between 16:00-17:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,333333333	L_q	0,166667 patients
P_1	0,333333333	W_q	0,027778 = 1,666667 min.
P_2	0,166666667	W_s	0,194444 = 11,66667 min.
P_3	0,166666667	L_s	1,166667 patients
P_4	0,083333333	ρ	0,500000
P_5	0,027777778		

As it is seen on the table, patients wait 1,666667 minutes between 16:00 and 17:00 on Monday that indicates that two doctors are enough for this period.

5.3.3.2 *Tuesday*

➤ *Time Period-1 (08:00-09:00)*

$\lambda = 25$ patients / hour

$\mu = 6$ patients / hour

$$\frac{25}{s * 6} \leq 1$$

There must be at least five doctors during the period. The results for $s=5$ are as in Table 5.30.

Table 5.30 Performance measures of the system between 08:00-09:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,009875998	L_q	0,516789 patients
P_1	0,04114999	W_q	0,020672 = 1,240294 min.
P_2	0,085729145	W_s	0,187338 = 11,24029 min.
P_3	0,119068258	L_s	4,683456 patients
P_4	0,124029435	ρ	0,833333
P_5	0,103357862		

On Tuesday between 08:00 and 09:00 five doctors are enough to keep patient waiting time acceptable.

➤ *Time Period-2 (09:00-10:00)*

$\lambda = 22$ patients / hour

$\mu = 6$ patients / hour

$$\frac{22}{s * 6} \leq 1$$

There must be at least four doctors during the period. The results for $s=4$ are as in Table 5.31.

Table 5.31 Performance measures of the system between 09:00-10:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,00909244	L_q	0,753265 patients
P_1	0,033338946	W_q	0,034239 = 2,054358 min.
P_2	0,061121401	W_s	0,200906 = 12,05436 min.
P_3	0,074703934	L_s	4,419931 patients
P_4	0,068478607	ρ	0,916667
P_5	0,251088224		

Patient waiting time, 2,054358 minutes, can be kept acceptable by four doctors between 09:00 and 10:00 on Tuesday.

➤ *Time Period-3 (10:00-11:00)*

$\lambda = 20$ patients / hour

$\mu = 6$ patients / hour

$$\frac{20}{s * 6} \leq 1$$

There must be at least four doctors during the period. The results for $s=4$ are as in Table 5.32.

Table 5.32 Performance measures of the system between 10:00-11:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,021310182	L_q	0,548101 patients
P_1	0,071033938	W_q	0,027405 = 1,644304 min.
P_2	0,118389897	W_s	0,194072 = 11,6443 min.
P_3	0,13154433	L_s	3,881435 patients
P_4	0,109620275	ρ	0,833333
P_5	0,365400918		

Patients wait average 1,644304 minutes between 10:00 and 11:00 on Tuesday so four doctors are enough for this period.

➤ *Time Period-4 (11:00-12:00)*

$\lambda = 14$ patients / hour

$\mu = 6$ patients / hour

$$\frac{14}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.33:

Table 5.33 Performance measures of the system between 11:00-12:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,064171123	L_q	0,47554 patients
P_1	0,14973262	W_q	0,033967 = 2,038027 min.
P_2	0,174688057	W_s	0,200634 = 12,03803 min.
P_3	0,135868489	L_s	2,808873 patients
P_4	0,158513237	ρ	0,777778
P_5	0,36986422		

As it is seen, three doctors are enough between 11:00 and 12:00 on Tuesday that patents wait 2,038027 minutes for consultation.

➤ *Time Period-5 (13:00-14:00)*

$\lambda = 11$ patients / hour

$\mu = 6$ patients / hour

$$\frac{11}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.34.

Table 5.34 Performance measures of the system between 13:00-14:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,043478261	L_q	0,803744 patients
P_1	0,079710145	W_q	0,073068 = 4,384058 min.
P_2	0,073067633	W_s	0,239734 = 14,38406 min.
P_3	0,133957327	L_s	2,637077 patients
P_4	0,122794216	ρ	0,916667
P_5	0,07504091		

Patient waiting time of 4,384058 minutes is longer but two doctors can be enough to meet the demand during this time period.

➤ *Time Period-6 (14:00-15:00)*

$\lambda = 15$ patients / hour

$\mu = 6$ patients / hour

$$\frac{15}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.35.

Table 5.35 Performance measures of the system between 14:00-15:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,04494382	L_q	0,585206 patients
P_1	0,112359551	W_q	0,039014 = 2,340824 min.
P_2	0,140449438	W_s	0,20568 = 12,34082 min.
P_3	0,117041199	L_s	3,085206 patients
P_4	0,146301498	ρ	0,833333
P_5	0,365753745		

As it is seen, three doctors are enough between 14:00 and 15:00 to keep patient waiting time between desired levels.

➤ *Time Period-7 (15:00-16:00)*

$\lambda = 8$ patients / hour

$\mu = 6$ patients / hour

$$\frac{8}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.36.

Table 5.36 Performance measures of the system between 15:00-16:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,2	L_q	0,355556 patients
P_1	0,266666667	W_q	0,044444 = 2,666667 min.
P_2	0,177777778	W_s	0,211111 = 12,66667 min.
P_3	0,237037037	L_s	1,688889 patients
P_4	0,158024691	ρ	0,666667
P_5	0,070233196		

Patient waiting time of 2,666667 minutes is acceptable so two doctors are enough to meet the demand between 15:00 and 16:00 on Tuesday.

➤ *Time Period-8 (16:00-17:00)*

$\lambda = 6$ patients / hour

$\mu = 6$ patients / hour

$$\frac{6}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=0$ are as in Table 5.37.

Table 5.37 Performance measures of the system between 16:00-17:00 on Tuesday

Probability	Value	Performance Measure	Value
P_0	0,333333333	L_q	0,166667 patients
P_1	0,333333333	W_q	0,027778 = 1,666667 min.
P_2	0,166666667	W_s	0,194444 = 11,66667 min.
P_3	0,166666667	L_s	1,166667 patients
P_4	0,083333333	ρ	0,500000
P_5	0,027777778		

As it is seen on the table, patients wait 1,666667 minutes for consultation so that two doctors will be enough on Tuesday between 16.00 and 17:00 to meet the demand while keeping patient waiting time at the desired level.

5.3.3.3 Wednesday

➤ Time Period-1 (08:00-09:00)

$$\lambda = 19 \text{ patients / hour}$$

$$\mu = 6 \text{ patients / hour}$$

$$\frac{19}{s * 6} \leq 1$$

There must be at least four doctors during the period. The results for $s=4$ are as in Table 5.38.

Table 5.38 Performance measures of the system between 08:00-09:00 on Wednesday

Probability	Value	Performance Measure	Value
P_0	0,028914888	L_q	0,460366 patients
P_1	0,091563814	W_q	0,02423 = 1,453787 min.
P_2	0,144976038	W_s	0,190896 = 11,45379 min.
P_3	0,153030262	L_s	3,627033 patients
P_4	0,121148958	ρ	0,791667
P_5	0,383638366		

Patient waiting time of 1,453787 minutes is too short so that four doctors will be enough on Wednesday between 08:00 and 09:00 to meet the demand while keeping patient waiting time under desired level.

➤ Time Period-2 (09:00-10:00)

$$\lambda = 17 \text{ patients / hour}$$

$$\mu = 6 \text{ patients / hour}$$

$$\frac{17}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results $s=3$ are as in Table 5.39.

Table 5.39 Performance measures of the system between 09:00-10:00 on Wednesday

Probability	Value	Performance Measure	Value
P_0	0,013143483	L_q	0,847035 patients
P_1	0,037239869	W_q	0,049826 = 2,989534 min.
P_2	0,05275648	W_s	0,216492 = 12,98953 min.
P_3	0,049825565	L_s	3,680368 patients
P_4	0,070586217	ρ	0,944444
P_5	0,199994281		

As it is seen, three doctors are enough between 09:00 and 10:00 on Wednesday to keep patient waiting time at the desired level.

➤ *Time Period-3 (10:00-11:00)*

$\lambda = 13$ patients / hour

$\mu = 6$ patients / hour

$$\frac{13}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.40.

Table 5.40 Performance measures of the system between 10:00-11:00 on Wednesday

Probability	Value	Performance Measure	Value
P_0	0,086083214	L_q	0,379417 patients
P_1	0,18651363	W_q	0,029186 = 1,751156 min.
P_2	0,202056432	W_s	0,195853 = 11,75116 min.
P_3	0,145929646	L_s	2,546084 patients
P_4	0,158090449	ρ	0,722222
P_5	0,342529307		

Patient waiting time of 1,751156 minutes is between limits so that three doctors will be enough on Wednesday between 10.00 and 11:00 to meet the demand.

➤ *Time Period-4 (11:00-12:00)*

$\lambda = 8$ patients / hour

$\mu = 6$ patients / hour

$$\frac{8}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.41.

Table 5.41 Performance measures of the system between 11:00-12:00 on Wednesday

Probability	Value	Performance Measure	Value
P_0	0,2	L_q	0,355556 patients
P_1	0,266666667	W_q	0,044444 = 2,666667 min.
P_2	0,177777778	W_s	0,211111 = 12,66667 min.
P_3	0,237037037	L_s	1,688889 patients
P_4	0,158024691	ρ	0,666667
P_5	0,070233196		

Patient waiting time can be kept at the desired level so two doctors are enough to meet the demand.

➤ *Time Period-5 (13:00-14:00)*

$\lambda = 7$ patients / hour

$\mu = 6$ patients / hour

$$\frac{7}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.42.

Table 5.42 Performance measures of the system between 13:00-14:00 on Monday

Probability	Value	Performance Measure	Value
P_0	0,263157895	L_q	0,250731 patients
P_1	0,307017544	W_q	0,035819 = 2,149123 min.
P_2	0,179093567	W_s	0,202485 = 12,14912 min.
P_3	0,208942495	L_s	1,417398 patients
P_4	0,121883122	ρ	0,583333
P_5	0,047398992		

Between 13:00 and 14:00 on Wednesday, patient waiting time, 2,149123 minutes, is short enough that two doctors can meet the demand.

➤ *Time Period-6 (14:00-15:00)*

$\lambda = 7$ patients / hour

$\mu = 6$ patients / hour

$$\frac{7}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.43.

Table 5.43 Performance measures of the system between 14:00-15:00 on Wednesday

Probability	Value	Performance Measure	Value
P_0	0,263157895	L_q	0,250731 patients
P_1	0,307017544	W_q	0,035819 = 2,149123 min.
P_2	0,179093567	W_s	0,202485 = 12,14912 min.
P_3	0,208942495	L_s	1,417398 patients
P_4	0,121883122	ρ	0,583333
P_5	0,047398992		

Also, between 13:00 and 14:00, patient waiting time is short enough to have two doctors working in the clinic.

➤ *Time Period-7 (15:00-16:00)*

$\lambda = 4$ patients / hour

$\mu = 6$ patients / hour

$$\frac{4}{s * 6} \leq 1$$

Only one doctor is enough during the period. The results for $s=1$ are as in Table 5.44.

Table 5.44 Performance measures of the system between 15:00-16:00 on Wednesday

Probability	Value	Performance Measure	Value
P_0	0,333333333	L_q	0,444444 patients
P_1	0,222222222	W_q	0,111111 = 6,666667 min.
P_2	0,148148148	W_s	0,277778 = 16,66667 min.
P_3	0,049382716	L_s	1,111111 patients
P_4	0,010973937	ρ	0,666667
P_5	0,001828989		

It is obvious that patient have longer waiting times such as 6,666667 minutes when there is only one doctor in the clinic but according to the desired level for waiting time of patients, one doctor is enough for this period.

➤ *Time Period-8 (16:00-17:00)*

$\lambda = 3$ patients / hour

$\mu = 6$ patients / hour

$$\frac{3}{s * 6} \leq 1$$

Only one doctor is enough during the period. The results for $s=1$ are as in Table 5.45.

Table 5.45 Performance measures of the system between 16:00-17:00 on Wednesday

Probability	Value	Performance Measure	Value
P_0	0,5	L_q	0,25 patients
P_1	0,25	W_q	0,083333 = 5 min.
P_2	0,125	W_s	0,25 = 15 min.
P_3	0,03125	L_s	0,75 patients
P_4	0,005208333	ρ	0,5
P_5	0,000651042		

Between 16:00 and 17:00 on Wednesday patients wait nearly half of the consultation time but this is acceptable so one doctor is enough in this period.

5.3.3.4 Thursday

➤ Time Period-1 (08:00-09:00)

$\lambda = 19$ patients / hour

$\mu = 6$ patients / hour

$$\frac{19}{s * 6} \leq 1$$

There must be at least four doctors during the period.. The results for $s=4$ are as in Table 5.46.

Table 5.46 Performance measures of the system between 08:00-09:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,028914888	L_q	0,460366 patients
P_1	0,091563814	W_q	0,02423 = 1,453787 min.
P_2	0,144976038	W_s	0,190896 = 11,45379 min.
P_3	0,153030262	L_s	3,627033 patients
P_4	0,121148958	ρ	0,791667
P_5	0,383638366		

As it is seen on the table, patient waiting time is too short so that four doctors will be enough on Thursday between 08:00 and 09:00 to meet the demand while keeping patient waiting time at the desired level.

➤ *Time Period-2 (09:00-10:00)*

$$\lambda = 14 \text{ patients / hour}$$

$$\mu = 6 \text{ patients / hour}$$

$$\frac{14}{s * 6} \leq 1$$

There must be at least three doctors during the period. Then “s” will be three and the other formulas mentioned in Chapter four are calculated. The results are as in Table 5.47.

Table 5.47 Performance measures of the system between 09:00-10:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,064171123	L_q	0,47554 patients
P_1	0,14973262	W_q	0,033967 = 2,038027 min.
P_2	0,174688057	W_s	0,200634 = 12,03803 min.
P_3	0,135868489	L_s	2,808873 patients
P_4	0,158513237	ρ	0,777778
P_5	0,36986422		

As it is seen, three doctors are enough between 09:00 and 10:00 that patients wait 2,038027 minutes for consultation.

➤ *Time Period-3 (10:00-11:00)*

$$\lambda = 16 \text{ patients / hour}$$

$$\mu = 6 \text{ patients / hour}$$

$$\frac{16}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for s=3 are as in Table 5.48.

Table 5.48 Performance measures of the system between 10:00-11:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,028037383	L_q	0,708896 patients
P_1	0,074766355	W_q	0,044306 = 2,658359 min.
P_2	0,099688474	W_s	0,210973 = 12,65836 min.
P_3	0,088611976	L_s	3,375562 patients
P_4	0,118149302	ρ	0,888889
P_5	0,315064805		

Patients wait 2,658359 minutes for consultation between 10:00 and 11:00 on Thursday but since it is acceptable for the management three doctors are enough for this period.

➤ *Time Period-4 (11:00-12:00)*

$\lambda = 10$ patients / hour

$\mu = 6$ patients / hour

$$\frac{7}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.49.

Table 5.49 Performance measures of the system between 11:00-12:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,090909091	L_q	0,631313 patients
P_1	0,151515152	W_q	0,063131 = 3,787879 min.
P_2	0,126262626	W_s	0,229798 = 13,78788 min.
P_3	0,21043771	L_s	2,29798 patients
P_4	0,175364759	ρ	0,833333
P_5	0,097424866		

Between 11:00 and 12:00, a waiting time of 3,787879 minutes occur for patients. It is a bit long but still acceptable that two doctors are enough.

➤ *Time Period-5 (13:00-14:00)*

$\lambda = 11$ patients / hour

$\mu = 6$ patients / hour

$$\frac{11}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.50.

Table 5.50 Performance measures of the system between 13:00-14:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,043478261	L_q	0,803744 patients
P_1	0,079710145	W_q	0,073068 = 4,384058 min.
P_2	0,073067633	W_s	0,239734 = 14,38406 min.
P_3	0,133957327	L_s	2,637077 patients
P_4	0,122794216	ρ	0,916667
P_5	0,07504091		

Patient waiting time is longer but acceptable between 13:00 and 14:00 so two doctors can be enough to meet the demand during this time period.

➤ *Time Period-6 (14:00-15:00)*

$\lambda = 12$ patients / hour

$\mu = 6$ patients / hour

$$\frac{12}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.51.

Table 5.51 Performance measures of the system between 14:00-15:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,111111111	L_q	0,296296 patients
P_1	0,222222222	W_q	0,024691= 1,481481 min.
P_2	0,222222222	W_s	0,191358= 11,48148 min.
P_3	0,148148148	L_s	2,296296 patients
P_4	0,148148148	ρ	0,666666
P_5	0,296296296		

Between 14:00 and 15:00 on Thursday, patient waiting time is short enough that three doctors can meet the demand while keeping patient waiting time at the desired level.

➤ *Time Period-7 (15:00-16:00)*

$\lambda = 7$ patients / hour

$\mu = 6$ patients / hour

$$\frac{7}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.52.

Table 5.52 Performance measures of the system between 15:00-16:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,263157895	L_q	0,250731 patients
P_1	0,307017544	W_q	0,035819 = 2,149123 min.
P_2	0,179093567	W_s	0,202485 = 12,14912 min.
P_3	0,208942495	L_s	1,417398 patients
P_4	0,121883122	ρ	0,583333
P_5	0,047398992		

Also, between 13:00 and 14:00, patient waiting time is short enough to have two doctors working in the clinic.

➤ *Time Period-8 (16:00-17:00)*

$\lambda = 4$ patients / hour

$\mu = 6$ patients / hour

$$\frac{4}{s * 6} \leq 1$$

Only one doctor is enough during the period. The results for $s=1$ are as in Table 5.53.

Table 5.53 Performance measures of the system between 16:00-17:00 on Thursday

Probability	Value	Performance Measure	Value
P_0	0,333333333	L_q	0,444444 patients
P_1	0,222222222	W_q	0,111111 = 6,666667 min.
P_2	0,148148148	W_s	0,277778 = 16,66667 min.
P_3	0,049382716	L_s	1,111111 patients
P_4	0,010973937	ρ	0,666667
P_5	0,001828989		

Patients wait more than the half of the consultation time between 16:00 and 17:00 on Thursday. It is because there is only one doctor in the clinic but such a waiting time is acceptable so it is not necessary to add another doctor for this period.

5.3.3.5 Friday

➤ *Time Period-1 (08:00-09:00)*

$\lambda = 13$ patients / hour

$\mu = 6$ patients / hour

$$\frac{13}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.54.

Table 5.54 Performance measures of the system between 08:00-09:00 on Friday

Probability	Value	Performance Measure	Value
P_0	0,086083214	L_q	0,379417 patients
P_1	0,18651363	W_q	0,029186 = 1,751156 min.
P_2	0,202056432	W_s	0,195853 = 11,75116 min.
P_3	0,145929646	L_s	2,546084 patients
P_4	0,158090449	ρ	0,722222
P_5	0,342529307		

As it is seen on the table, patient waiting time is between limits so that three doctors will be enough on Friday between 08.00 and 09:00 to meet the demand.

➤ *Time Period-2 (09:00-10:00)*

$\lambda = 8$ patients / hour

$\mu = 6$ patients / hour

$$\frac{8}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.55.

Table 5.55 Performance measures of the system between 09:00-10:00 on Friday

Probability	Value	Performance Measure	Value
P_0	0,2	L_q	0,355556 patients
P_1	0,266666667	W_q	0,044444 = 2,666667 min.
P_2	0,177777778	W_s	0,211111 = 12,66667 min.
P_3	0,237037037	L_s	1,688889 patients
P_4	0,158024691	ρ	0,666667
P_5	0,070233196		

Patient waiting time of 2,666667 minutes is acceptable so two doctors are enough to meet the demand during this period.

➤ *Time Period-3 (10:00-11:00)*

$\lambda = 9$ patients / hour

$\mu = 6$ patients / hour

$$\frac{9}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.56.

Table 5.56 Performance measures of the system between 10:00-11:00 on Friday

Probability	Value	Performance Measure	Value
P_0	0,142857143	L_q	0,482143 patients
P_1	0,214285714	W_q	0,053571= 3,214286min.
P_2	0,160714286	W_s	0,220238= 13,21429min.
P_3	0,241071429	L_s	1,982143 patients
P_4	0,180803571	ρ	0,750000
P_5	0,090401786		

Patient waiting time between 10:00 and 11:00 is a bit longer than the previous time periods but it is still acceptable so two doctors can meet the demand.

➤ *Time Period-4 (11:00-12:00)*

$\lambda = 8$ patients / hour

$\mu = 6$ patients / hour

$$\frac{8}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.57.

Table 5.57 Performance measures of the system between 11:00-12:00 on Friday

Probability	Value	Performance Measure	Value
P_0	0,2	L_q	0,355556 patients
P_1	0,266666667	W_q	0,044444 = 2,666667 min.
P_2	0,177777778	W_s	0,211111 = 12,66667 min.
P_3	0,237037037	L_s	1,688889 patients
P_4	0,158024691	ρ	0,666667
P_5	0,070233196		

The patient arrival rate during this period is the same as the second period so two doctors are enough for also this period.

➤ *Time Period-5 (13:00-14:00)*

$\lambda = 15$ patients / hour

$\mu = 6$ patients / hour

$$\frac{15}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.58.

Table 5.58 Performance measures of the system between 13:00-14:00 on Friday

Probability	Value	Performance Measure	Value
P_0	0,04494382	L_q	0,585206 patients
P_1	0,112359551	W_q	0,039014 = 2,340824 min.
P_2	0,140449438	W_s	0,20568 = 12,34082 min.
P_3	0,117041199	L_s	3,085206 patients
P_4	0,146301498	ρ	0,833333
P_5	0,365753745		

As it is seen, three doctors are enough between 13:00 and 14:00 that patients wait 2,340824 minutes for consultation.

➤ *Time Period-6 (14:00-15:00)*

$\lambda = 13$ patients / hour

$\mu = 6$ patients / hour

$$\frac{13}{s * 6} \leq 1$$

There must be at least three doctors during the period. The results for $s=3$ are as in Table 5.59.

Table 5.59 Performance measures of the system between 14:00-15:00 on Friday

Probability	Value	Performance Measure	Value
P_0	0,086083214	L_q	0,379417 patients
P_1	0,18651363	W_q	0,029186 = 1,751156 min.
P_2	0,202056432	W_s	0,195853 = 11,75116 min.
P_3	0,145929646	L_s	2,546084 patients
P_4	0,158090449	ρ	0,722222
P_5	0,342529307		

As it is seen on the table, patient waiting time is acceptable so that three doctors will be enough during this period to meet the demand while keeping patient waiting time under desired level.

➤ *Time Period-7 (15:00-16:00)*

$\lambda = 9$ patients / hour

$\mu = 6$ patients / hour

$$\frac{9}{s * 6} \leq 1$$

There must be at least two doctors during the period. The results for $s=2$ are as in Table 5.60.

Table 5.60 Performance measures of the system between 15:00-16:00 on Friday

Probability	Value	Performance Measure	Value
P_0	0,142857143	L_q	0,482143 patients
P_1	0,214285714	W_q	0,053571= 3,214286min.
P_2	0,160714286	W_s	0,220238= 13,21429min.
P_3	0,241071429	L_s	1,982143 patients
P_4	0,180803571	ρ	0,750000
P_5	0,090401786		

Patient waiting time can be kept at the desired limit so two doctors are enough for this period to meet the demand.

➤ *Time Period-8 (16:00-17:00)*

$\lambda = 5$ patients / hour

$\mu = 6$ patients / hour

$$\frac{5}{s * 6} \leq 1$$

Only one doctor is enough during the period. The results for $s=1$ are as in Table 5.61.

Table 5.61 Performance measures of the system between 16:00-17:00 on Friday for $s=1$

Probability	Value	Performance Measure	Value
P_0	0,166666667	L_q	0,694444 patients
P_1	0,138888889	W_q	0,138889 = 8,333333 min.
P_2	0,115740741	W_s	0,305556 = 18,33333 min.
P_3	0,048225309	L_s	1,527778 patients
P_4	0,013395919	ρ	0,833333
P_5	0,002790816		

During this period patients have longest waiting time of 8,333333 minutes that is nearly the same as the consultation time. Although this waiting time can be acceptable it can be examined adding another doctor worth to implement or not.

For $s=2$:

Table 5.62 Performance measures of the system between 16:00-17:00 on Friday for $s=2$

Probability	Value	Performance Measure	Value
P_0	0,411764706	L_q	0,245098 patients
P_1	0,343137255	W_q	0,04902 = 2,941176 min.
P_2	0,285947712	W_s	0,215686 = 12,94118 min.
P_3	0,11914488	L_s	1,078431 patients
P_4	0,0330958	ρ	0,416667
P_5	0,006894958		

As it is seen, if two doctors are assigned to this period, patient waiting time decreases from 8,333 minutes to 2,941 minutes. The management can decide which assignment is worth to apply.

Here, the doctor requirement between 16:00 and 17:00 on Friday is determined as two doctors because there is not an additional cost of assigning another doctor in this study.

As a result, doctor requirements for all periods of each day are as follows:

Table 5.63 Doctor Requirements per period and per day

	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	TOTAL
08:00-09:00	5	5	4	4	3	21
09:00-10:00	4	4	3	3	2	16
10:00-11:00	4	4	3	3	2	16
11:00-12:00	3	3	2	2	2	12
13:00-14:00	3	2	2	2	3	12
14:00-15:00	3	3	2	2	3	13
15:00-16:00	2	2	1	2	2	9
16:00-17:00	2	2	1	1	1	7
TOTAL	26	25	18	19	18	106

5.3.4 *Personnel Scheduling Models*

After determining doctor requirements per period and per day, now it is necessary to schedule the working patterns of these doctors. For solving this scheduling problem, formulas given in Chapter four are used and an integrated scheduling model is built by different shift lengths per day. After determining required work patterns and required shifts for doctors, a simple heuristic is used to assign start times to the optimum work patterns.

Since the number of residents in the department is known, limited staff size version of the models described previously will be used. The doctors are assumed to work minimum fifteen hours a week and three different shifts of 3-, 4- and 5-hours a day are considered. If a doctor works three hours, four hours and five hours a day, then he / she works five days, four days and three days a week respectively. It is also assumed that the days and periods doctors work must be consecutive.

The clinic session starts at 08:00 and finishes at 17:00. Then, there are eight time periods with equal length of one hour. A doctor may start working at the beginning of any period according to his / her shift pattern. Therefore the model includes 8-hour a day and alternative 3-, 4- and 5-hour shift as mentioned above.

The objective of the integrated model is to minimize total number of weekly labor and understaffing level. According to the limited staff size version of the problem Day-Of-Week problem is solved first. Then the results of the DOW problem are used as the constraints of Time-Of-Day problem.

The general form of the DOW scheduling problem for the limited staff size is as follows:

$$\min Z = \sum_{j=1}^J X_j + 100 \sum_{i=1}^7 U_i$$

subject to

$$\sum_{j=1}^J A_{ij} X_j - O_i + U_i = D_i \quad i = 1, 2, \dots, 7$$

$$\sum_{j=1}^J X_j = 24$$

$$X_j \geq 0 \quad \text{and integer}$$

where D_i values are the total values in the last row of Table 5.63.

The objective of the model is minimizing total work force and understaffing. The coefficient of understaffing is assigned as 100 because it is desired to take understaffing level low. The first constraint gives the optimum number of doctors working for each work pattern in order to meet the demand. To take a feasible solution, understaffing and overstaffing variables are added to the first constraint. The second constraint limits the total number of workforce since there are 24 residents in the department.

The A_{ij} matrix with 3-, 4- and 5- hour shifts can be seen in Table 5.64:

Table 5.64 A_{ij} matrix for the DOW problem

	3-hr	4-hr		5-hr		
Days / Patterns	1	2	3	4	5	6
MONDAY	1	1	0	1	0	0
TUESDAY	1	1	1	1	1	0
WEDNESDAY	1	1	1	1	1	1
THURSDAY	1	1	1	0	1	1
FRIDAY	1	0	1	0	0	1

The model is solved by Lingo 8.0 and the solution report of the DOW problem can be seen in Appendix F. The results indicate that eighteen doctors should work Monday through Friday with 3-hour shift type. Two doctors should work Monday

through Thursday with 4-hour shift type. Four doctors should work Monday through Wednesday with 5-hour shift type.

The last column of the solution report for variables indicates the reduced cost of the variable. The reduced cost indicates the objective coefficient of the variable would have to improve before it would take place in the solution. As it is seen in the solution report, the reduced costs of the variables that have nonzero values are zero since they take place in the solution. But the other variables whose values are zero have a reduced cost of 1. This means that if the objective coefficients of these variables are increased by 1, they can enter the solution.

The slack or surplus column in the solution report tells how close the solution to satisfy a constraint. The constraints with slack or surplus values of 0 are satisfied completely.

The solution report also gives a dual price figure for each constraint. Dual price indicates the amount by which the objective function will improve given a unit of increase in the right hand side of the constraint. For example, for the first constraint (row 2), if the right hand side of the constraint increases by 1, the amount of improvement in the objective function will be -1. Since the objective function is minimization, then the objective value will increase by 1.

The results of the DOW problem are summarized below:

Table 5.65 Limited staff size for work patterns for doctors

Shift Length	Days on Patterns	Req. Number
3-Hour	M-T-W-Th-F	18
4-Hour	M-T-W-Th	2
5-Hour	M-T-W	4

While the 3-, 4- and 5-hour shift patterns are given, the start time of each shift type is not given. TOD problem is solved to determine the start time of each shift type on each day. The general form of the TOD problem is as follows:

$$\min Z_i = \sum_{k=1}^{K_i} X_{ik} + *100 \sum_{l=1}^{L_i} U_{il}$$

subject to

$$\sum_{k=1}^{K_i} A_{ik} X_{ik} - O_{il} + U_{il} = D_{il} \quad l = 1, 2, \dots, L_i$$

$$\sum_{k \in 3} X_{ik} = W_3$$

$$\sum_{k \in 4} X_{ik} = W_4$$

$$\sum_{k \in 5} X_{ik} = W_5$$

$$X_{ik} \geq 0 \quad \text{and integer}$$

W_3 , W_4 and W_5 values indicate required workforce size for 3-, 4- and 5-hour shift types respectively.

The objective of the function is again minimizing total workforce scheduled and total understaffing. The first constraint gives the optimum number of doctors working for each shift pattern in order to meet the demand. To obtain a feasible solution, understaffing and overstaffing variables are added to the constraint. The second, third and the fourth constraints provide that the number of doctors scheduled for each shift type is equal to the required numbers calculated by the DOW problem.

Then A_{ij} matrix with the 3-, 4- and 5-hours shift combinations are as in Table 5.66:

Table 5.66 A_{ij} matrix for the TOD problem

	3-Hour						4-Hour					5-Hour			
Period / Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
08:00-09:00	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0
09:00-10:00	1	1	0	0	0	0	1	1	0	0	0	1	1	0	0
10:00-11:00	1	1	1	0	0	0	1	1	1	0	0	1	1	1	0
11:00-12:00	0	1	1	1	0	0	1	1	1	1	0	1	1	1	1
12:00-13:00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13:00-14:00	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1
14:00-15:00	0	0	0	1	1	1	0	0	1	1	1	0	1	1	1
15:00-16:00	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1
16:00-17:00	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1

By using the A_{ij} matrix above, the TOD problem is solved for each day separately.

The solution reports of the TOD problem for each day can be seen in Appendix G-1 through G-5. The interpretation of reduced cost, slack or surplus and dual price columns are as in the DOW problem.

The results of the TOD problem can be summarized as below:

Table 5.67 Limited staff size for required shifts for doctors

Start Time	MONDAY			TUESDAY			WEDNESDAY			THURSDAY			FRIDAY		
	3-Hr	4-Hr	5-Hr	3-Hr	4-Hr	5-Hr	3-Hr	4-Hr	5-Hr	3-Hr	4-Hr	5-Hr	3-Hr	4-Hr	5-Hr
08:00	-	2	3	-	2	3	-	2	2	4	-	-	3	-	-
11:00	-	-	1	-	-	1	-	-	2	-	2	-	2	-	-
13:00	-	-	-	-	-	-	-	-	-	13	-	-	1	-	-
14:00	18	-	-	18	-	-	18	-	-	1	-	-	12	-	-

This table shows that on Monday 18 doctors start to work at 14:00 with a 3-hour shift and they finish working at 17:00. Two doctors start to work at 08:00 with a 4-hour shift and finish working at 12:00. With the 5-hour shift, three doctors start to work at 08:00 and finish working 14:00 while one doctor works between 09:00 and 15:00.

Then, the results for integrated solution for the personnel scheduling problem in DEU, Dept. of Ort. and Trau. are as follows:

Table 5.68 Limited staff size integrated model's solution

TOD	Model's Assignments				
	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY
08:00-09:00	5	5	4	4	3
09:00-10:00	0	0	0	0	0
10:00-11:00	0	0	0	0	0
11:00-12:00	1	1	2	2	2
12:00-13:00	0	0	0	0	0
13:00-14:00	0	0	0	13	1
14:00-15:00	18	18	18	1	12
15:00-16:00	0	0	0	0	0
16:00-17:00	0	0	0	0	0
Total	24	24	24	20	18

The table shows that on Monday and Tuesday five doctors start to work at 08:00, one doctor start at 11:00 and 18 doctors start to work at 14:00. On Wednesday, four doctors start to work at 08:00 while two doctors start at 11:00. Again eighteen doctors start to work at 14:00. Thursday is different from Wednesday that thirteen doctors start to work at 13:00 and one doctor starts at 14:00. On Friday, three doctors start to work at 08:00 and two doctors start at 11:00. One of the rest starts at 13:00 while thirteen doctors start the clinic at 14:00. Consequently, it is seen that optimal shifts start at 08:00, 11:00, 13:00 and 14:00. The model does not assign shifts beginning at the other periods.

Now, it is necessary to combine the seven TOD problems solutions with the work patterns obtained from the DOW problems solutions and to assign start times for the days of each optimum work pattern for doctors. The assignment table can be seen at Table 5.69:

Table 5.69 Assignment table for doctors

Shift Type	Start time	Requirement
3-hr	M ₁₄	18
	T ₁₄	18
	W ₁₄	18
	Th ₈	4
	Th ₁₃	13
	Th ₁₄	1
	F ₈	3
	F ₁₁	2
	F ₁₃	1
	F ₁₄	12
4-hr	M ₈	2
	T ₈	2
	W ₈	2
	Th ₁₁	2
5-hr	M ₈	3
	M ₁₁	1
	T ₈	3
	T ₁₁	1
	W ₈	2
	W ₁₁	2

For the first doctor, the first pattern at the top of the list at Table 5.65, which is 3-hour from Monday through Friday, is selected and required number for that pattern is decreased by. Then, 3-hour shifts at Table 5.67 are reduced by one on each required day. The assignment for the other doctors continues in the same manner until all doctors are assigned to a pattern and shift. The assignment for 24 doctors in the department is as in Table 5.70:

Table 5.70 Required work patterns with their start times

Doctor	Shift	Start Time				
		M	T	W	Th	F
1	3	14:00	14:00	14:00	14:00	14:00
2	3	14:00	14:00	14:00	13:00	14:00
3	3	14:00	14:00	14:00	13:00	14:00
4	3	14:00	14:00	14:00	13:00	14:00
5	3	14:00	14:00	14:00	13:00	14:00
6	3	14:00	14:00	14:00	13:00	14:00
7	3	14:00	14:00	14:00	13:00	14:00
8	3	14:00	14:00	14:00	13:00	14:00
9	3	14:00	14:00	14:00	13:00	14:00
10	3	14:00	14:00	14:00	13:00	14:00
11	3	14:00	14:00	14:00	13:00	14:00
12	3	14:00	14:00	14:00	13:00	14:00
13	3	14:00	14:00	14:00	13:00	13:00
14	3	14:00	14:00	14:00	13:00	11:00
15	3	14:00	14:00	14:00	08:00	11:00
16	3	14:00	14:00	14:00	08:00	08:00
17	3	14:00	14:00	14:00	08:00	08:00
18	3	14:00	14:00	14:00	08:00	08:00
19	4	08:00	08:00	08:00	11:00	-
20	4	08:00	08:00	08:00	11:00	-
21	5	08:00	08:00	08:00	-	-
22	5	08:00	08:00	08:00	-	-
23	5	08:00	08:00	11:00	-	-
24	5	11:00	11:00	11:00	-	-

Now, this table shows that which doctor starts to work on which period of which day. The second column indicates the shift type assigned to the doctor while the rest indicates the optimum start times for the doctor. According to the Table 5.70, two doctors, doctors 19 and 20, works 16 hours a week while the others work 15 days. This problem can be solved by assigning a priority rule among the doctors according to their seniority. The two doctors with the least seniority can be given the 4-hour shift type with the work pattern of Monday through Thursday. Another way to deal with this problem is making rotations among doctors. For example, the doctors, working 4-hr shift one week, can work 5-hr or 3-hr shift next week. By this way shifts can be distributed among doctors fairly.

CHAPTER SIX

CONCLUSION

This chapter summarizes the research performed in this thesis including the appendices following. Also directions for future research are proposed in the latter sections of the chapter.

6.1 Performed Research

As it is mentioned in the first chapter, the main interest of the study is scheduling working hours of doctors in DEU, Dept.of Ort. and Trau. The doctors in this department have different working patterns that some works on an appointment based system and others work in outpatient clinic. Then, the work schedule of the doctors working by appointment is determined by using appointment scheduling approach and the work schedule of doctors working in the outpatient clinic is obtained by using personnel scheduling approach.

In Chapter five, approaches for solving both appointment scheduling and personnel scheduling problems are presented. The application in DEU, Dept.of Ort. and Trau. is given. First appointment scheduling problem is solved for one doctor. Model parameters are determined according to the data gathered. Two different versions for this problem are evaluated: dynamic and static state. For the dynamic state, dynamic programming approach is used and two alternative models are developed with different number of time periods in a clinic session. The first alternative model is solved for four time periods and the f table set is obtained that is used for scheduling patients throughout the clinic session. For the second alternative model, calculations are made for three time periods and again f table set is obtained. Since dynamic programming approach has a disadvantage that the decision of a stage cannot be taken unless the previous stage is finished, a more efficient solution is obtained by converting dynamic programming calculations into static state. Branch-and-bound approach is used for the static version and again two alternative models are developed as in the dynamic state. The optimal patient assignment for the first

alternative model with four time periods is (2,1,1,2) which means that for the beginning of the clinic session, two patients must be scheduled. For the second and third period, only one patient should be scheduled in order to minimize patient waiting time and doctor idle time. Since all of the remaining patients must be scheduled in the last period, two patients must be at the beginning of the last period.

Then, the personnel scheduling problem in DEU, Dept.of Ort. and Trau. is solved. The first step of the solution is determining number of doctors required per period by using queuing theory. To determine the parameters of queuing theory input analysis is performed by the data gathered from the patient records. After determining the model parameters, performance measures of the system are calculated by using the queuing formulas. The basic performance measure that is used to determine number of doctors required is patient waiting time. Optimal doctor requirements for each period are determined by keeping patient waiting time acceptable. Once these requirements are calculated, linear programming approach is used to schedule working hours of the doctors. The limited staff size version of the integrated model given in Chapter four is used to solve the scheduling problem. First DOW problem is solved to determine the optimum work patterns for the doctors. Then, the TOD problem is solved for each day separately by using the results of the DOW problem. After determining optimal work and shift patterns for doctors, these patterns are assigned to each doctor by a simple heuristic.

6.2 Future Research

There is a great deal of research to be made on appointment scheduling. The proposed solution to the problem of determining optimum number of appointments per time period can be improved by further research. While the appointment system considered in this thesis based on a real-world system, there are still some assumptions unrealistic and future research should attempt to remove them. For instance, all patients and the doctor are assumed to be punctual and it is suggested for further research to evaluate the effects of earliness and lateness on the optimal schedule. The other features that are suggested for further research are preemption of

unscheduled patients such as walk-ins and emergencies, no-show probabilities of patients and server failures. These are important considerations in some applications and are not uncommon in real-world systems but they have not been incorporated to the current solution.

The further research suggested about personnel scheduling is to take the special requests of the doctors into consideration. Some of the doctors may want to work or not to work on particular days. New assignment models can be developed by using models developed in this thesis in order to satisfy these requests. Also, priorities among the doctors can be given among the doctors can be given so that the requests of these doctors must be satisfied firstly. In this study, all doctors are assumed to be identical in terms of their qualifications but doctors can be classified and evaluated according to their qualifications as future research.

REFERENCES

- Arena 5.0 (2000), Rockwell Software Inc.
- Avriel, M. & Golany, B.(1996). *Mathematical Programming for Industrial Engineers*.New York: Marcel Dekker.
- Babes, M. & Sarma, G.V. (1991). Outpatient Queues at the Ibn-Rochd Health Center. *Journal of the Operational Research Society*, 42, 845-855.
- Bailey, N. (1952). A Study of Queues and Appointment Systems in Hospital outpatient Departments with Special Reference to Waiting Times. *Journal of the Royal Statistical Society*, 14, 185-199.
- Banks, J. & Gibson, R.R. (1997). Don't Simulate When: 10 Rules for Determining When Simulation is not Appropriate. *IIE Solutions*, September.
- Banks, J., Carson, J.S., Nelson, B. & Nicol, D.M. (2001). *Discrete-Event System Simulation*. (3rd Edition). New York: Prentice Hall.
- Brahimi, M. & Worthington D.J. (1991-a). The Finite Capacity Multi-Server Queue with Inhomogeneous arrival Rate and Discrete Service Time Distribution and its Application to Continuous Service Time Problems. *European Journal of Operational Research*, 50, 310-324
- Brahimi, M. & Worthington D.J. (1991-b). Queuing Models for Outpatient Appointment Systems: A Case Study. *Journal of the Operational Research Society*, 42, 733-746.
- Cayırlı, Tugba & Veral, Emre (2003). Outpatient Scheduling in health Care : A Review of Literature. *Working Paper*
- Cox, T.F., Birchall, J.F. & Wong, H. (1985). Optimizing the Queuing System for an Ear, Nose and Throat Outpatient Clinic. *Journal of Applied Statistics*, 12,113-126.

- Denton, B. & Gupta, D. (2001). A Sequential Bounding Approach for Optimal Appointment Scheduling. *Unpublished Working paper, DeGroote School of Business, McMaster University, Hamilton, Ontario.*
- Fries, B. & Marathe V. (1981). Determination of Optimal Variable-Sized Multiple-Block Appointment Systems. *Operations Research*, 29, 324-345.
- Gorantla, B. R. (1995). *Heuristic Optimization and Improvement of Personnel Scheduling Problem in Nurse Scheduling Domain*, Dissertation. The University of Texas at Arlington.
- Harper, P.R. & Gamlin, H.M. (2003). Reduced Outpatient Waiting Times with Improved Appointment Scheduling: A Simulation Modeling Approach. *OR Spectrum*, 25, 205-222.
- Hillier, F.S. & Lieberman, G.J. (1995). *Introduction to Operations Research*. (6th Edition). New York: Mc-Graw Hill.
- Ho, C. & Lau, H. (1992). Minimizing Total Cost in Scheduling Outpatient Appointments. *Management Science*, 38, 1750-1764.
- Ho, C. & Lau, H. (1995). Introducing Variable-Interval Appointment Scheduling in Service Systems. *International Journal of Production & Operations Management*, 15, 59-69.
- Ho, C. & Lau, H. (1999). Evaluating the Impact of Operating Conditions on the Performance of Appointment Scheduling Rules in Service Systems. *European Journal of Operational Research*, 112, 542-553.
- Huarng, F. & Lee, M.H. (1996). Using Simulation in Outpatient Queues: A Case Study. *International Journal of Health Care Quality Assurance*, 9 (6), 21-25.
- Jansson, B. (1966). Choosing a Good Appointment System-A Study of Queues of the Type (D/M/1). *Operations Research*, 14, 292-312.

- Katz, J. (1969). Simulation of Outpatient Appointment Systems. *Communications of the ACM*, 12, 215-222.
- Kelton, W.D., Sadowski, R.P. & Sadowski, D.A.(2002). *Simulation with Arena*. (2nd Edition). New York: McGraw Hill.
- Klassen, K.J. & Rohleder, T. R. (1996). Scheduling Outpatient Appointments in a Dynamic Environment. *Journal of Operations Management*, 14, 83-101.
- Krajewski, L.J. & Ritzman, L.P (2002). *Operations Management: Strategy and Analysis*. (6th Edition). Prentice Hall.
- Lapin, L.L. (1991). *Quantitative Methods for Business Decisions*. (5th Edition).Florida: Harcourt Brace Jovanovich Inc.
- Lehaney, B., Clarke, S.A. & Paul, R.J. (1999). A Case of Intervention in an Outpatinets Department. *Journal of the Operational Research Society*, 50, 877-891.
- Liao, C., Pegden, C.D. & Rosenshine, M. (1993). Planning Timely Arrivals to a Stochastic Production or Service System. *Journal of Operational Research Society*, 50, 877-891.
- Liu, L. & Liu, X. (1998). Dynamic and Static Job Allocation for Multi-Server Systems. *IIE Transactions*, 30, 845-854.
- Liu, L. & Liu, X. (1998). Block Appointment Systems for Outpatient Clinics with Multiple Doctors. *Journal of the Operational Research Society*, 49, 1254-1259.
- Lingo Release 8.0 (2003), Lindo Systems Inc.
- Mahachek, A.R. & Knabe, T.L. (1984). Computer Simulation of Patient Flow in Obstetrical/Gynecology Clinics. *Simulation*, 43(August), 95-101.

- Mondschein, S.V. & Weintraub, G.Y. (2003). Appointment Policies in Service Operations: A Critical Analysis of the Economic Framework. *Production and Operations Management*, 12(2), 266-286.
- O'Keefe, R. (1985). Investigating Outpatient Departments: Implementable Policies and Qualitative Approaches. *Journal of Operational Research Society*, 36, 705-712.
- Pegden, D. C., Shannon, R.E. & Sadowski, R.P. (1995). *Introduction to Simulation Using SIMAN*. New York: Mc-Graw Hill.
- Pinedo, M. & Chao, X. (1999). *Operations Scheduling with Applications in Manufacturing and Services*. McGraw Hill.
- Rising, E., Baron, R. & Averill, B. (1973). A System Analysis of a University Health Service Outpatient Clinic. *Operations Research*, 21, 1030-1047.
- Robinson, L.W. & Chen, R.R. (2001). Scheduling Doctors' Appointments: Optimal and Empirically-Based Heuristic Policies. *Unpublished Working Paper*, Johnson graduate School of Management, Cornell University, Ithaca, New York.
- Rohleder, T.R. & Klassen, K.J. (2001). Using Client-Variance Information to Improve Dynamic Appointment Scheduling Performance. *Omega*, 28, 239-302.
- Soriano, A. (1966). Comparison of Two Scheduling Systems. *Operations Research*, 14, 388-397.
- Su, S. & Shih, C. (2003). Managing a Mixed-Registration-Type Appointment System in Outpatient Clinics. *International Journal of Medical Informatics*, 70, 31-40.
- Swisher, J.R., Jacobson, S.H., Jun, J.B. & Balci, O. (2001). Modeling and Analyzing a Physician Clinic Environment Using Discrete-Event Simulation. *Computers & Operations Research*, 28, 105-125.
- Vanden Bosch, P.M., Dietz, C.D. and Simeoni J.R. (1999). Scheduling Customer Arrivals to a Stochastic Service System. *Naval Research Logistics*, 46, 549-559.

- Vanden Bosch, P.M., Dietz, C.D. (2000). Minimizing Expected Waiting in a Medical Appointment System. *IIE Transactions*, 32, 841,848.
- Visser, J. (1979). Selecting a Suitable Appointment System in an Outpatient Setting. *Medical Care*, 17 (12), 1207-1220.
- Wang, P.P. (1993). Static and Dynamic Scheduling of Customer Arrivals to a Single-Server System. *Naval Research Logistics*, 40, 345-360.
- Wang, P.P. (1997). Optimally Scheduling N Customer Arrival Times for a Single-Server System. *Computer & Operations Research*, 24, 703-716.
- White, B. & Pike, M.C. (1964). Appointment Systems in Outpatients' Clinics and the Effect of Patients' Unpunctuality.
- Yang, K. K., Lu, M.L. & Quek, S.A. (1998). A New Appointment Rule for a Single-Server, Multiple-Customer Service System. *Naval Research Logistics*, 45, 313-326.

Patient Arrival Data on 21.04.2005

	DATE	HOOR	PATIENT FILE	NAME-SURNAME	USER		
1	21.04.2005	08:20	253991	KAMER KILIC	BELMA CELIK	SAGLIK MD.	
2	21.04.2005	09:15	120533	FATMA PENBE AKINCI	BELMA CELIK	EMEKLİ SANDIGI	
3	21.04.2005	09:24	290087	MERVE YASAR	BELMA CELIK	HASTANELER	
4	21.04.2005	10:20	524112	NIL DEMIRCAN	BELMA CELIK	DOKUZ EYLUL UNV.	
5	21.04.2005	11:06	907079	MURUVVET OZGULER	SEDA AYAZ	EMEKLİ SANDIGI	
6	21.04.2005	11:36	64556	EMETI BENLI	OZCAN ERTEN	EMEKLİ SANDIGI	
7	21.04.2005	11:40	986872	EMRE OZAY	BELMA CELIK	MILLI EGITIM	
8	21.04.2005	14:16	103169	AYSE YUCE	BELMA CELIK	DOKUZ EYLUL HAST.	
9	21.04.2005	14:35	1001259	ISMAIL GOKSU	SEDA AYAZ	EMNIYET MD.	
10	21.04.2005	15:04	688992	ZUBEYDE BABACAN	BELMA CELIK	OZEL SAHIS	
11	21.04.2005	15:33	13753	SAMI GUNTAS	BELMA CELIK	DOKUZ EYLUL REKTORLUK	
12	21.04.2005	16:04	797001	KARDELEN AYDOGDU	BELMA CELIK	DOKUZ EYLUL HAST.	
13	21.04.2005	16:39	809451	ERDAL ASLAN	BELMA CELIK	DOKUZ EYLUL UNV.	
14	21.04.2005	16:41	856416	SIMAY POLATTURK	BELMA CELIK	DOKUZ EYLUL REKTORLUK	
15	21.04.2005	16:56	357605	HASAN CULLU	BELMA CELIK	DOKUZ EYLUL HAST.	

APPENDIX B

TOTAL PATIENT ARRIVALS

Total Number of Patient Arrivals per Period and per Day

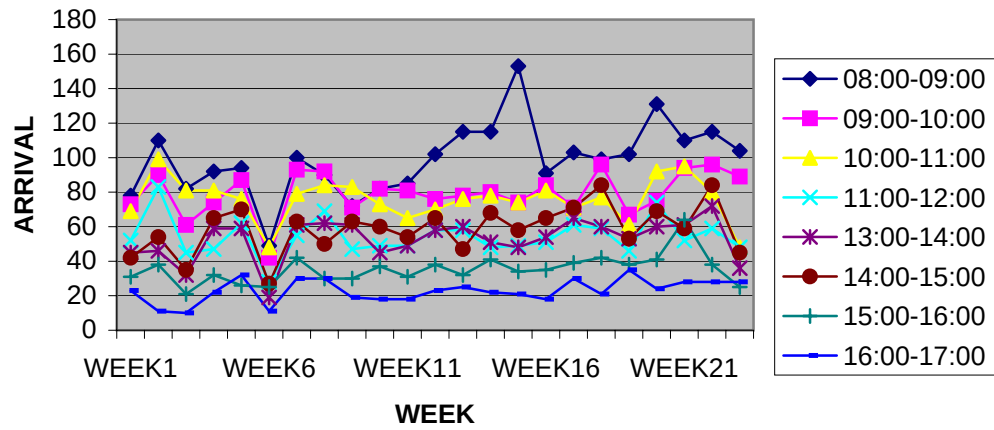
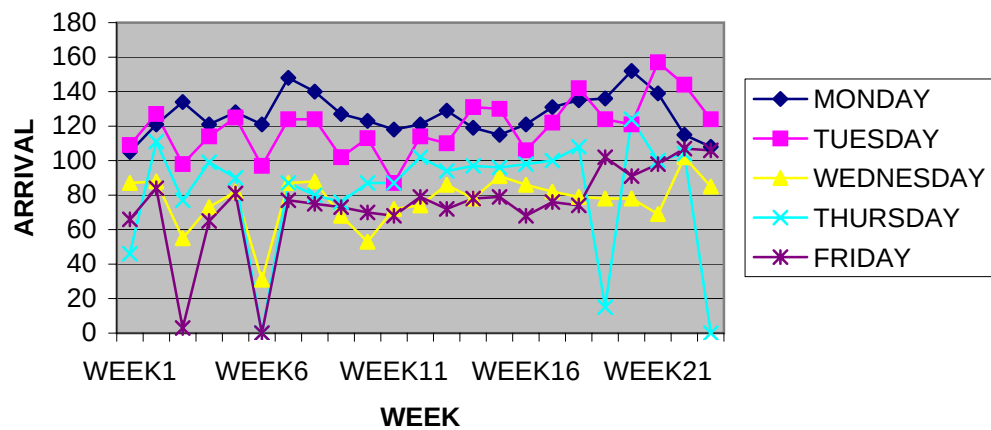
	08:00-09:00	09:00-10:00	10:00-11:00	11:00-12:00	13:00-14:00	14:00-15:00	15:00-16:00	16:00-17:00	TOTAL
13.12.2004	18	27	17	17	7	10	2	7	105
14.12.2004	22	10	18	17	14	12	9	7	109
15.12.2004	15	22	17	6	10	7	7	3	87
16.12.2004	11	9	9	4	3	3	5	2	46
17.12.2004	12	5	8	8	11	10	8	4	66
20.12.2004	30	24	23	14	9	13	5	3	121
21.12.2004	23	23	22	21	9	12	13	4	127
22.12.2004	18	16	16	19	6	8	5	0	88
23.12.2004	21	17	25	20	9	9	7	3	111
24.12.2004	18	10	13	9	13	12	8	1	84
27.12.2004	29	26	24	9	13	18	9	6	134
28.12.2004	21	16	19	16	5	7	11	3	98
29.12.2004	14	7	18	7	5	4	0	0	55
30.12.2004	18	12	19	11	9	6	1	1	77
31.12.2004	0	0	1	2	0	0	0	0	3
03.01.2005	22	19	17	16	17	11	10	9	121
04.01.2005	20	16	23	9	14	24	6	2	114
05.01.2005	16	16	18	4	4	8	3	4	73
06.01.2005	20	14	18	11	13	14	7	2	99
07.01.2005	14	9	5	7	11	8	6	5	65
10.01.2005	30	23	25	12	11	9	7	11	128
11.01.2005	20	28	14	19	13	20	5	6	125
12.01.2005	20	15	10	13	3	14	5	2	82
13.01.2005	9	14	16	11	13	12	8	7	90
14.01.2005	15	7	11	7	19	15	1	6	81
17.01.2005	23	14	26	18	9	12	11	8	121
18.01.2005	18	16	15	6	10	15	14	3	97
19.01.2005	8	12	7	4	0	0	0	0	31
20.01.2005	0	0	0	0	0	0	0	0	0
21.01.2005	0	0	0	0	0	0	0	0	0
24.01.2005	27	19	26	18	19	19	11	9	148
25.01.2005	29	19	28	17	7	12	6	6	124
26.01.2005	23	29	14	5	4	5	3	4	87
27.01.2005	13	6	7	13	21	14	10	3	87
28.01.2005	8	20	4	2	10	13	12	8	77
31.01.2005	37	20	21	18	12	11	11	10	140
01.02.2005	20	32	22	15	7	12	4	12	124
02.02.2005	12	22	18	15	12	4	4	1	88
03.02.2005	11	12	16	8	11	13	5	4	80

Total Number of Patient Arrivals per Period and per Day (Cont'd)

	08:00-09:00	09:00-10:00	10:00-11:00	11:00-12:00	13:00-14:00	14:00-15:00	15:00-16:00	16:00-17:00	TOTAL
04.02.2005	10	6	7	13	20	10	6	3	75
07.02.2005	20	21	28	14	15	19	8	2	127
08.02.2005	24	18	18	9	10	12	4	7	102
09.02.2005	11	11	13	6	10	8	5	4	68
10.02.2005	10	11	14	10	11	10	6	4	76
11.02.2005	7	10	10	8	15	14	7	2	73
14.02.2005	23	27	26	11	5	21	6	4	123
15.02.2005	20	14	22	14	9	16	13	5	113
16.02.2005	14	10	4	5	9	4	5	2	53
17.02.2005	20	23	10	8	7	6	11	2	87
18.02.2005	5	8	11	11	15	13	2	5	70
21.02.2005	25	21	19	20	12	10	5	6	118
22.02.2005	18	20	19	9	9	7	3	2	87
23.02.2005	16	16	7	7	8	5	7	6	72
24.02.2005	18	18	12	9	8	10	8	4	87
25.02.2005	8	6	8	4	12	22	8	0	68
28.02.2005	35	15	17	15	9	14	9	7	121
01.03.2005	22	24	16	14	9	13	8	8	114
02.03.2005	14	15	15	10	6	11	3	0	74
03.03.2005	19	17	15	14	12	14	7	4	102
04.03.2005	12	5	7	5	22	13	11	4	79
07.03.2005	24	18	21	25	13	12	10	6	129
08.03.2005	30	23	17	9	10	13	6	2	110
09.03.2005	23	17	13	11	6	6	6	4	86
10.03.2005	26	16	17	2	18	6	4	5	94
11.03.2005	12	4	8	11	13	10	6	8	72
14.03.2005	30	14	22	11	13	18	7	4	119
15.03.2005	29	27	20	13	14	13	7	8	131
16.03.2005	23	10	13	8	8	12	2	2	78
17.03.2005	18	22	16	8	7	12	10	4	97
18.03.2005	15	7	7	8	9	13	15	4	78
21.03.2005	37	15	20	10	9	13	7	4	115
22.03.2005	40	19	23	14	9	10	9	6	130
23.03.2005	30	20	17	6	8	6	2	2	91
24.03.2005	34	8	9	12	9	12	6	6	96
25.03.2005	12	12	5	7	13	17	10	3	79
28.03.2005	24	17	23	15	13	13	14	2	121
29.03.2005	19	27	14	11	7	18	8	2	106
30.03.2005	20	22	11	12	7	6	5	3	86
31.03.2005	18	12	22	7	15	15	4	5	98
01.04.2005	10	6	11	6	12	13	4	6	68
04.04.2005	32	20	16	17	14	19	8	5	131
05.04.2005	22	22	23	17	11	16	7	4	122
06.04.2005	18	8	12	10	12	8	6	8	82

Total Number of Patient Arrivals per Period and per Day (Cont'd)

	08:00-09:00	09:00-10:00	10:00-11:00	11:00-12:00	13:00-14:00	14:00-15:00	15:00-16:00	16:00-17:00	TOTAL
07.04.2005	15	17	15	12	13	15	8	5	100
08.04.2005	16	4	5	5	15	13	10	8	76
11.04.2005	22	23	26	19	11	22	10	2	135
12.04.2005	27	30	23	13	15	20	9	5	142
13.04.2005	21	25	10	5	6	5	2	5	79
14.04.2005	20	14	16	9	17	24	4	4	108
15.04.2005	9	4	2	13	11	13	17	5	74
18.04.2005	27	27	15	17	12	14	11	13	136
19.04.2005	30	14	19	13	15	21	6	6	124
20.04.2005	23	14	13	5	4	6	9	4	78
21.04.2005	1	2	1	3	0	2	2	4	15
22.04.2005	21	10	13	8	22	10	10	8	102
25.04.2005	37	17	22	21	20	19	14	2	152
26.04.2005	29	16	21	23	9	13	5	5	121
27.04.2005	22	19	9	5	5	10	5	3	78
28.04.2005	30	12	27	19	11	11	6	8	124
29.04.2005	13	11	13	6	15	16	11	6	91
02.05.2005	35	20	21	18	14	18	11	2	139
03.05.2005	22	38	32	10	11	16	14	14	157
04.05.2005	16	9	12	7	13	4	6	2	69
05.05.2005	21	17	20	8	8	11	14	1	100
06.05.2005	16	10	10	9	15	10	19	9	98
09.05.2005	16	22	14	21	15	15	8	4	115
10.05.2005	27	29	25	14	15	18	9	7	144
11.05.2005	28	16	21	9	7	13	6	2	102
12.05.2005	26	18	12	9	11	13	9	6	104
13.05.2005	18	11	8	6	24	25	6	9	107
16.05.2005	33	28	12	15	2	8	5	5	108
17.05.2005	33	25	14	10	13	15	6	8	124
18.05.2005	22	25	8	8	3	12	2	5	85
19.05.2005	0	0	0	0	0	0	0	0	0
20.05.2005	16	11	14	15	18	10	12	10	106

WEEKLY ARRIVALS / PERIOD**WEEKLY ARRIVALS / DAY**

Patient Arrivals per Period on Friday

	08:00-09:00	09:00-10:00	10:00-11:00	11:00-12:00	13:00-14:00	14:00-15:00	15:00-16:00	16:00-17:00	TOTAL
17.12.2004	12	5	8	8	11	10	8	4	66
24.12.2004	18	10	13	9	13	12	8	1	84
31.12.2004	0	0	1	2	0	0	0	0	3
07.01.2005	14	9	5	7	11	8	6	5	65
14.01.2005	15	7	11	7	19	15	1	6	81
21.01.2005	0	0	0	0	0	0	0	0	0
28.01.2005	8	20	4	2	10	13	12	8	77
04.02.2005	10	6	7	13	20	10	6	3	75
11.02.2005	7	10	10	8	15	14	7	2	73
18.02.2005	5	8	11	11	15	13	2	5	70
25.02.2005	8	6	8	4	12	22	8	0	68
04.03.2005	12	5	7	5	22	13	11	4	79
11.03.2005	12	4	8	11	13	10	6	8	72
18.03.2005	15	7	7	8	9	13	15	4	78
25.03.2005	12	12	5	7	13	17	10	3	79
01.04.2005	10	6	11	6	12	13	4	6	68
08.04.2005	16	4	5	5	15	13	10	8	76
15.04.2005	9	4	2	13	11	13	17	5	74
22.04.2005	21	10	13	8	22	10	10	8	102
29.04.2005	13	11	13	6	15	16	11	6	91
06.05.2005	16	10	10	9	15	10	19	9	98
13.05.2005	18	11	8	6	24	25	6	9	107
20.05.2005	16	11	14	15	18	10	12	10	106

APPENDIX E

RESULTS OF INPUT ANALYSIS

193

Results of Input Analysis via Arena 5.0

		SAMPLE MEAN	SAMPLE STD. DEV.	SQUARE ERROR	NUMBER OF INTERVALS	DEGREES OF FREEDOM	TEST STATISTICS	P- VALUE
MONDAY	08:00-09:00	28	6,27	0,046007	3	1	1,94	0,183
	09:00-10:00	21	4,37	0,02328	3	1	0,119	0,738
	10:00-11:00	21	4,41	0,035533	4	2	3,89	0,158
	11:00-12:00	16	3,91	0,027649	3	1	1,56	0,225
	13:00-14:00	12	4,16	0,038007	4	2	3,01	0,23
	14:00-15:00	15	4,12	0,04303	3	1	1,57	0,223
	15:00-16:00	9	2,95	0,038609	4	2	0,812	0,677
	16:00-17:00	6	3,15	0,038868	3	1	2,52	0,12
TUESDAY	08:00-09:00	25	5,57	0,05607	4	2	2,63	0,276
	09:00-10:00	22	6,82	0,049069	4	2	3,17	0,217
	10:00-11:00	20	4,52	0,029821	4	2	4,14	0,138
	11:00-12:00	14	4,22	0,048855	3	1	0,361	0,567
	13:00-14:00	11	2,74	0,059128	3	1	5,84	0,0174
	14:00-15:00	15	4,25	0,059959	3	1	0,342	0,58
	15:00-16:00	8	3,25	0,038459	3	1	2,03	0,173
	16:00-17:00	6	3,06	0,024427	3	1	0,218	0,668
WEDNESDAY	08:00-09:00	19	5,39	0,051547	4	2	3,67	0,176
	09:00-10:00	17	5,71	0,053112	3	1	2,17	0,158
	10:00-11:00	13	4,28	0,031733	3	1	1,62	0,218
	11:00-12:00	8	3,06	0,026098	3	1	1,49	0,232
	13:00-14:00	7	3,23	0,020745	3	1	0,447	0,506
	14:00-15:00	7	3,46	0,044083	3	1	1,34	0,248
	15:00-16:00	4	2,28	0,050973	3	1	0,647	0,445
	16:00-17:00	3	2,05	0,042336	4	2	1,57	0,467
THURSDAY	08:00-09:00	19	6,57	0,066962	3	1	0,134	0,728
	09:00-10:00	14	4,37	0,065766	3	1	2,86	0,0933
	10:00-11:00	16	5,23	0,040437	3	1	0,926	0,364
	11:00-12:00	10	4,25	0,032776	3	1	0,593	0,46
	13:00-14:00	11	4,22	0,046702	3	1	1,51	0,23
	14:00-15:00	12	4,47	0,041249	3	1	0,409	0,532
	15:00-16:00	7	2,94	0,015894	3	1	0,0018	0,75
	16:00-17:00	4	1,84	0,029809	3	1	0,134	0,728
FRIDAY	08:00-09:00	13	4,1	0,042427	3	1	0,96	0,355
	09:00-10:00	8	3,75	0,036375	3	1	2,5	0,122
	10:00-11:00	9	3,33	0,05592	3	1	2,29	0,145
	11:00-12:00	8	3,19	0,024716	3	1	1,26	0,268
	13:00-14:00	15	4,17	0,073529	3	1	2,15	0,16
	14:00-15:00	13	4,08	0,124063	3	1	2,96	0,0889
	15:00-16:00	9	4,53	0,049759	3	1	1,73	0,206
	16:00-17:00	5	2,77	0,033409	3	1	0,434	0,515

APPENDIX F

SOLUTION REPORT OF THE DOW PROBLEM

Global optimal solution found at iteration: 4
 Objective value: 324.00000

Variable	Value	Reduced Cost
X1	18,0000	0,0000
X2	2,0000	0,0000
X3	0,0000	1,0000
X4	4,0000	0,0000
X5	0,0000	1,0000
X6	0,0000	2,0000
U1	2,0000	0,0000
U2	1,0000	0,0000
U3	0,0000	1,0000
U4	0,0000	1,0000
U5	0,0000	1,0000

Row	Slack or Surplus	Dual price
1	27,0000	-1,0000
2	0,0000	-1,0000
3	0,0000	-1,0000
4	0,0000	0,0000
5	0,0000	0,0000
6	0,0000	0,0000
7	0,0000	1,0000

APPENDIX G-1

SOLUTION REPORT OF THE TOD PROBLEM FOR MONDAY

Global optimal solution found at iteration: 10

Objective value: 24.000

Variable	Value	Reduced Cost
X1	0,0000	0,0000
X2	0,0000	0,0000
X3	0,0000	0,0000
X4	0,0000	0,0000
X5	0,0000	0,0000
X6	18,0000	0,0000
X7	2,0000	0,0000
X8	0,0000	0,0000
X9	0,0000	0,0000
X10	0,0000	0,0000
X11	0,0000	0,0000
X12	3,0000	0,0000
X13	0,0000	0,0000
X14	0,0000	0,0000
X15	1,0000	0,0000
U1	0,0000	100,0000
U2	0,0000	100,0000
U3	0,0000	100,0000
U4	0,0000	100,0000
U5	0,0000	100,0000
U6	0,0000	100,0000
U7	0,0000	100,0000
U8	0,0000	100,0000

Row	Slack or Surplus	Dual Price
1	24,0000	-1,0000
2	0,0000	0,0000
3	0,0000	0,0000
4	0,0000	0,0000
5	0,0000	0,0000
6	0,0000	0,0000
7	0,0000	0,0000
8	0,0000	0,0000
9	0,0000	0,0000
10	0,0000	-1,0000
11	0,0000	-1,0000
12	0,0000	-1,0000

APPENDIX G-2

SOLUTION REPORT OF THE TOD PROBLEM FOR TUESDAY

Global optimal solution found at iteration: 10

Objective value: 24.000

Variable	Value	Reduced Cost
X1	0,0000	0,0000
X2	0,0000	0,0000
X3	0,0000	0,0000
X4	0,0000	0,0000
X5	0,0000	0,0000
X6	18,0000	0,0000
X7	2,0000	0,0000
X8	0,0000	0,0000
X9	0,0000	0,0000
X10	0,0000	0,0000
X11	0,0000	0,0000
X12	3,0000	0,0000
X13	0,0000	0,0000
X14	0,0000	0,0000
X15	1,0000	0,0000
U1	0,0000	100,0000
U2	0,0000	100,0000
U3	0,0000	100,0000
U4	0,0000	100,0000
U5	0,0000	100,0000
U6	0,0000	100,0000
U7	0,0000	100,0000
U8	0,0000	100,0000

Row	Slack or Surplus	Dual Price
1	24,0000	-1,0000
2	0,0000	0,0000
3	0,0000	0,0000
4	0,0000	0,0000
5	0,0000	0,0000
6	0,0000	0,0000
7	0,0000	0,0000
8	0,0000	0,0000
9	0,0000	0,0000
10	0,0000	-1,0000
11	0,0000	-1,0000
12	0,0000	-1,0000

APPENDIX G-3

SOLUTION REPORT OF THE TOD PROBLEM FOR WEDNESDAY

Global optimal solution found at iteration: 10

Objective value: 24.00000

Variable	Value	Reduced Cost
X1	0,0000	0,0000
X2	0,0000	0,0000
X3	0,0000	0,0000
X4	0,0000	0,0000
X5	0,0000	0,0000
X6	18,0000	0,0000
X7	2,0000	0,0000
X8	0,0000	0,0000
X9	0,0000	0,0000
X10	0,0000	0,0000
X11	0,0000	0,0000
X12	2,0000	0,0000
X13	0,0000	0,0000
X14	0,0000	0,0000
X15	2,0000	0,0000
U1	0,0000	100,0000
U2	0,0000	100,0000
U3	0,0000	100,0000
U4	0,0000	100,0000
U5	0,0000	100,0000
U6	0,0000	100,0000
U7	0,0000	100,0000
U8	0,0000	100,0000

Row	Slack or Surplus	Dual Price
1	24,0000	-1,0000
2	0,0000	0,0000
3	0,0000	0,0000
4	0,0000	0,0000
5	0,0000	0,0000
6	0,0000	0,0000
7	0,0000	0,0000
8	0,0000	0,0000
9	0,0000	0,0000
10	0,0000	-1,0000
11	0,0000	-1,0000
12	0,0000	-1,0000

APPENDIX G-4

SOLUTION REPORT OF THE TOD PROBLEM FOR THURSDAY

Global optimal solution found at iteration: 11

Objective value: 20.00000

Variable	Value	Reduced Cost
X1	4,0000	0,0000
X2	0,0000	0,0000
X3	0,0000	0,0000
X4	0,0000	0,0000
X5	13,0000	0,0000
X6	1,0000	0,0000
X7	0,0000	0,0000
X8	0,0000	0,0000
X9	0,0000	0,0000
X10	2,0000	0,0000
X11	0,0000	0,0000
X12	0,0000	1,0000
X13	0,0000	1,0000
X14	0,0000	1,0000
X15	0,0000	1,0000
U1	0,0000	100,0000
U2	0,0000	100,0000
U3	0,0000	100,0000
U4	0,0000	100,0000
U5	0,0000	100,0000
U6	0,0000	100,0000
U7	0,0000	100,0000
U8	0,0000	100,0000

Row	Slack or Surplus	Dual Price
1	20,0000	-1,0000
2	0,0000	0,0000
3	0,0000	0,0000
4	0,0000	0,0000
5	0,0000	0,0000
6	0,0000	0,0000
7	0,0000	0,0000
8	0,0000	0,0000
9	0,0000	0,0000
10	0,0000	-1,0000
11	0,0000	-1,0000
12	0,0000	0,0000

APPENDIX G-5

SOLUTION REPORT OF THE TOD PROBLEM FOR FRIDAY

Global optimal solution found at iteration: 9

Objective value: 18.00000

Variable	Value	Reduced Cost
X1	3,0000	0,0000
X2	0,0000	0,0000
X3	0,0000	0,0000
X4	2,0000	0,0000
X5	1,0000	0,0000
X6	12,0000	0,0000
X7	0,0000	1,0000
X8	0,0000	1,0000
X9	0,0000	1,0000
X10	0,0000	1,0000
X11	0,0000	1,0000
X12	0,0000	1,0000
X13	0,0000	1,0000
X14	0,0000	1,0000
X15	0,0000	1,0000
U1	0,0000	100,0000
U2	0,0000	100,0000
U3	0,0000	100,0000
U4	0,0000	100,0000
U5	0,0000	100,0000
U6	0,0000	100,0000
U7	0,0000	100,0000
U8	0,0000	100,0000

Row	Slack or Surplus	Dual Price
1	18,0000	-1,0000
2	0,0000	0,0000
3	0,0000	0,0000
4	0,0000	0,0000
5	0,0000	0,0000
6	0,0000	0,0000
7	0,0000	0,0000
8	0,0000	0,0000
9	0,0000	0,0000
10	0,0000	-1,0000
11	0,0000	0,0000
12	0,0000	0,0000

APPENDIX H

LINGO

LINGO is a simple tool for performing complex and powerful tasks. It is easy to use with only forty-plus commands. It enables the user to formulate large problems concisely, solve them and analyze the solution. At the simplest level, with LINGO, equations with many independent variables can be solved by entering just a few simple lines. At the most complex level, it means that LINGO can generate very large and complicated optimization models, with many functions, in comparatively few lines of input.

The maximum size problem LINGO can handle depends on the version. Specifically, the limits are as follows:

Version	Maximum Rows	Maximum Columns	Memory Required
Regular/Student	10	200	3 Mb
Super	500	1000	4 Mb
Hyper	2000	4000	6 Mb
Industrial	8000	16000	8 Mb
Extended	32000	100000	16 Mb

The version of the Lingo used in this thesis is the extended version.

LINGO is a mathematical modeling language which lets the user express the problem in a natural manner which is very similar to standard mathematical notation. This means that rather than entering each term of each constraint explicitly, a whole series of similar constraints can be expressed in a single compact statement. Also, LINGO allows the user to express the model using conventional mathematical terms which are not allowed in many other packages such as subscripted variables, sets and operations over sets (summation etc.).

Three types of models are solved by using LINGO. The first one is the direct models in which all variable values are fixed and may be determined directly without resorting to a solver. With direct models, LINGO functions like a spreadsheet or calculator. The second model solved by LINGO is simultaneous models that consist of interdependent equations which cannot be solved entirely by direct subscription. LINGO uses the appropriate solver, linear or nonlinear; to find a set of variable values that simultaneously satisfies all the interdependent equations. The last model LINGO solves is the optimization model. The models developed in this thesis are optimization models. Models can be optimized by specifying which variables a model should be minimized or maximized for the most desirable result. When an optimization model is run, LINGO presents the solution for which the selected variables have the minimum or maximum value.

The models solved by LINGO have some other properties. Firstly, LINGO lets the user to do What-If analysis of models by specifying input parameters each time you solve the model. Then, set-based models can be developed such as in this thesis. Set-based models are useful for problems involving large numbers of similar variables and constraints. Most larger real-world problems can be expressed efficiently using set-based models. Also, LINGO allows the user to restrict variables to take only integer values as in the models developed in this thesis.

LINGO has four solvers: a direct solver, a simultaneous linear solver/optimizer, a simultaneous nonlinear solver/optimizer and a branch-and-bound manager for models with integer restrictions. The personnel scheduling models developed in this thesis are integer restricted so branch-and-bound approach is used by LINGO to solve the problems.