Journal of Service Science and Management, 2019, 12, 589-604
http://www.scirp.org/journal/jssm
ISSN Online: 1940-9907
ISSN Print: 1940-9893

On Performance of Prioritized Appointment Scheduling for Healthcare

Yang Guo, Yong Yao

Blekinge Institute of Technology, Karlskrona, Sweden
Email: yang.guo@bth.se, yong.yao@bth.se

How to cite this paper: Guo, Y. and Yao, Y. (2019) On Performance of Prioritized Appointment Scheduling for Healthcare. *Journal of Service Science and Management*, **12**, 589-604.
https://doi.org/10.4236/jssm.2019.125040

Received: May 23, 2019
Accepted: August 2, 2019
Published: August 5, 2019

Abstract

Designing the appointment scheduling is a challenging task for the development of healthcare system. The efficient solution approach can provide high-quality healthcare service between care providers (CP)s and care receivers (CR)s. In this paper, we consider the healthcare system with the heterogeneous CRs in terms of urgent and routine CRs. Our suggested model assumes that the system gives the service priority to the urgent CRs by allowing them to interrupt the ongoing routine appointments. An appointment handoff scheme is suggested for the interrupted routine appointments, and thus the routine CRs can attempt to re-establish the appointment scheduling with other available CPs. With these considerations, we study the scheduling performance of the system by using the Markov chains based modeling approach. The numerical analysis is reported and the simulation experiment is conducted to validate the numerical results.

Keywords

Appointment Scheduling, Healthcare System, Queueing Theory, Discrete Time Markov Chain (DTMC)

1. Introduction

Today's demand for healthcare is dramatically increasing as the factor of the aging population and expectations growing during the past few years [1] [2]. This raises the need to develop the substantial healthcare services with innovative technologies from both industry and academia [3] [4]. A major technical challenge refers to the requirement for the provision of high-quality healthcare services delivered from the Care Provider (CP) like, e.g., a nurse, a doctor, to the Care Receiver (CR) such as a patient [5]. One important research focus associated with this challenge is laid on the problem called appointment scheduling

[6] [7]. This is because the suitable solution to the appointment scheduling can enhance the efficiency of healthcare delivery, and thus improving the quality of healthcare services.

Designing the efficient appointment scheduling is a complicated process due to the crucial responsibility of dealing with the limited resources (e.g., in terms of hardware, people, time availabilities) for the concurrent timely-access by multiple users (*i.e.*, CRs). Connected to this process, the interactivities between CPs and CRs need to be modeling, and needs to be measured and optimized. This can be performed by using different methods such as analytical approaches, simulation experiments and practical measurements on the healthcare system [8] [9] [10]. Most of previous studies done along with this research line mainly consider the statistical characterizations of timely-access activities of CRs. The corresponding examples are the average waiting time of CRs for the new appointment and the average service time of CPs dealing with the secluded appointment.

Apart from the above-addressed statistical characterizations, another important factor affecting the performance of appointment scheduling is related to the heterogeneous aspect of the healthcare system. In our work, such heterogeneous aspect is expressed in the form of the diversity of CRs, which inherently exists in the healthcare system. A typical example is that the CRs consist of both routine and urgent patients. The urgent patients are usually given a higher priority over the routine ones to have the appointments. Under this situation, the scheduled appointments for routine CRs may be interrupted. As a result, the appointment scheduling performance of routine CRs may degrade.

In this paper, the motivation comes down to the numerical analysis on the prioritized appointment scheduling for the healthcare system. To do this, a Discrete Time Markov Chain (DTMC) based queueing model is built up to theoretically represents the appointment scheduling in the presence of both routine and urgent CRs. Moreover, a new scheme called *appointment handoff* is suggested for routine CRs to deal with their interruption by urgent CRs. To evaluate the performance of the suggested scheme, we use metrics in terms of the probabilities of blocking and terminating the appointments of urgent CRs, the appointment-completion throughput and the average service time.

The rest of the paper is as follows: Section 2 presents the background related to the appointment scheduling and the urgent patient modeling, together with our contribution. Section 3 describes the system model used in our study. A DTMC based queueing model is built up in Section 4. The numerical and simulation results for performance evaluation are discussed in Section 5. Finally, the paper is concluded in Section 6.

2. Related Work

A large number of studies on appointment scheduling have been reported in recent literature. For instance, the authors of [11] suggest an appointment scheduling solution by considering the patient-provider mutual preference. This solution is used to reduce the schedule fragmentation, and thus yielding a higher appointment acceptance rate and clinic time utilization rate. In [12], the authors suggest a load balancing strategy for aid scheduling in eHealth. The goal is to optimize the Quality of Service (QoS) of healthcare delivery. In [13], the authors suggest a metaheuristic approach to solve the appointment scheduling problem. This approach is developed based on the Greedy algorithm and Tabu search mechanism.

Further, the particular interest associated with the healthcare appointment scheduling is placed on the system performance evaluation. To do this, two typical methodologies are widely used in many studies. They are the numerical analysis and the simulation or practical implementation based experiments.

The work done on numerical analysis of the performance of carrying out appointment scheduling is as follows. In [14], the authors consider the appointment scheduling problem for patients visiting a given place. A queueing theory model is further built up and the corresponding numerical analysis is reported. In [15], the authors report on a mathematical model to study the queue dynamics of blood collection system. The goal is to use the simulation-optimization based approach to improve the performance of the blood donor's appointment scheduling. In [16], the paper considers the problem of how to optimize the number of patient appointment. To solve the problem, the authors develop a stochastic mathematical overbooking model. The goal is to maximize the expected total profits for diverse healthcare environments. In [9], the authors consider the no-show behaviour of patients, who have successfully booked the appointments. Based on the single server M/D/1/K queueing approach, they develop a social welfare function that makes the trade-off between patients reward and cost. As such, the numerical analysis is conducted to study the impact of no-show rate on the outcome of social welfare.

Regarding the simulation or experimental based performance evaluation, the corresponding studies are reported below. In [17], the authors develop a new appointment scheduling system. This system takes into account both patients and providers preferences. The simulation-based performance evaluation is conducted to show the effectiveness of the developed system. The main advantage of such solution is to provide patients with more flexibility when they are involved in the scheduling process. Similar studies can be found in [18] [19]. In [8] and [20], the optimization on patients' waiting time in healthcare services especially with appointment scheduling is studied. The simulation-based experiments are conducted to find the suitable solution. In [21], the paper reports the problem of surgery appointment scheduling with the limited resources in terms of operating rooms and surgery durations. The authors address that such problem is presumably NP-hard. To tackle the computational complexity, the authors suggest a set of algorithms together with simulation experiments. In [22], the authors report on the development of a web-technology based solution for doctors to handle the appointment scheduling with patients Similarly, the authors of [23] and [24] report on an intelligent appointment scheduling system, which can be deployed in hospital with using the Near Field Communication

(NFC).

Moreover, the variety of CRs existent in the healthcare system is also widely investigated. In [25], the authors suggest a Markov Decision Process (MDP) based model to schedule six different types of patients. The goal of this model is to maximize the average revenue of accepting outpatients and minimize the average overtime penalty. In [26], the authors suggest a simulation model for doing appointment scheduling between CPs and CRs. They consider two types of patients (*i.e.*, the new and existing patients) for further simulation-based study. The goal of this simulation model is to find the best balance between new and existing patients with respect to their arriving time and average service-completion time. In [27], the authors consider the problem of scheduling patients to visit a cancer infusion room with limited resources in terms of chairs and nurses. The authors further suggest the acuity-based rules to schedule two different types of patients, *i.e.*, the high-acuity and low-acuity based patients. In [28], the authors consider the problem regarding the outpatient appointment schedule problem with routine and urgent patients. The authors develop a numerical solution approach under the assumption of deterministic service time together with the no-show behaviour of patients.

The work reported above has laid the ground to investigate the effect of heterogeneous aspect of CRs (e.g., patients) on the appointment scheduling in the healthcare system. However, to the best of our knowledge, there are few studies done so far on dealing with the interactions among different types of CRs, together with the corresponding feedback activities. In this paper, our main contribution is to suggest an appointment handoff scheme for the CRs to deal with the appointment interruption. We also build up a DTMC based queueing model for the numerical analysis purposes. We also carry out the simulation based performance evaluation. Both numerical and simulation results show the feasibility of the suggested scheme.

3. System Model

A particular eHealth system is considered in our study. Different parameters are used in the system modeling as indicated in **Table 1**. The detailed descriptions are as follows.

3.1. Care Provider Model

In the system, there are M CPs, which are denoted by $p_1, p_2, \cdots, p_m, \cdots, p_M$, respectively. These CPs provide the candidate appointments to CRs during the prescribed time period. As shown in **Figure 1**, this time period is assumed to consist of multiple identical time slots. These time slots are denoted by a set with infinite elements, *i.e.*, $T = \{t_k \mid k = 0,1,2,\cdots\}$. Each time slot has a uniform value δ, which can be equal to, like, e.g., 20 minutes, 1 hour. Further, two neighbouring time slots may not be continuous along the time domain, for instance, t_k = [10:00 AM, 11:00 AM] and t_{k+1} = [2:00 PM, 3:00 PM].

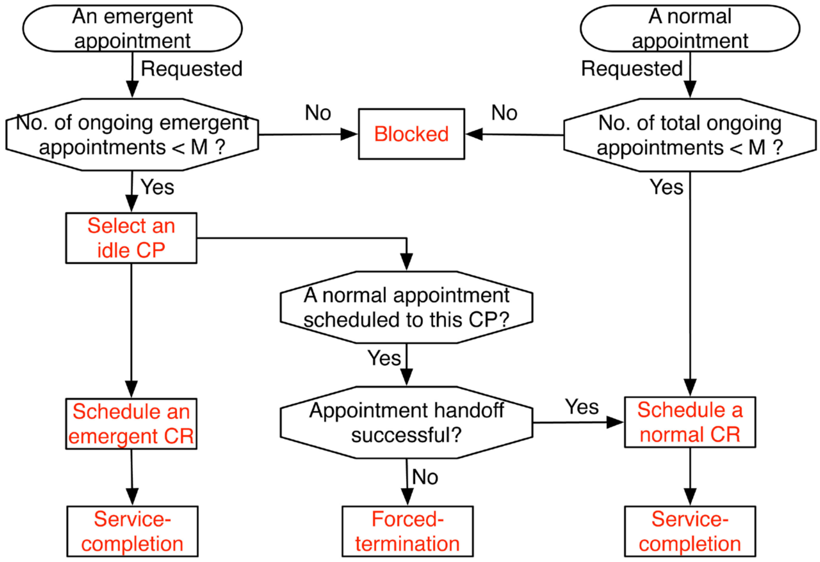

Figure 1. Appointment scheduling schemes.

Table 1. Parameter Notations.

Parameter Notations	Definition
M	Number of CPs in the system.
(i, j)	A pair of values indicating a system state that i ongoing routine appointments and j ongoing urgent appointments.
S	System state space.
$\pi_{i,j}$	Steady-state probability of Markov state (i, j).
λ_1	Mean request rate of appointments by urgent CRs.
λ_2	Mean request rate of appointments by routine CRs.
μ_1	Mean service-completion rate of ongoing urgent appointments for a single CP.
μ_2	Mean service-completion rate of ongoing routine appointments for a single CP.
λ_j^*	Mean rate of allocating CPs to the appointment requests by urgent CRs, depending on the value j.
μ_j^*	Mean rate of completing the appointments by urgent CRs, depending on the value j.
$\lambda_{i,j}^{\dagger}$	Mean rate of allocating CPs to the appointment requests by routine CRs, depending on the values of i and j
$\mu_{i,j}^{\dagger}$	Mean rate of completing the appointments by routine CRs, depending on the values of i and j.

During each time slot, each CP is assumed to be able to only deal with a single appointment with a single CR. Let two states *free* and *busy* denote the appointment status of a particular CP. Here, the state *free* refers to the event that there is no booked appointment and the particular CP is available for a new appointment demanded by a CR. Similarly, the state *busy* refers to the event that there exists an appointment booked for the particular CP and a CR.

3.2. Care Receiver Model

We assume that there are two different types of CRs existing in the system. They are called *urgent* CRs and *routine* CRs, respectively. Both urgent and routine CRs need the appointments with CPs.

The arrivals of the appointment requested by CRs are assumed to independently follow the Poisson process with mean rates λ_1 and λ_2 for urgent and routine CRs, respectively. The time periods of the appointments between CPs and CRs are assumed to be exponentially distributed with average values $1/\mu_1$ and $1/\mu_2$ for urgent and routine CRs, respectively.

Further, the two pairs of values $\{1/\lambda_1, 1/\mu_1\}$ and $\{1/\lambda_2, 1/\mu_2\}$ are actually equivalent to the average time periods of the two states *free* and *busy* for urgent and routine CRs, respectively. Because the two states may cross one or more consecutive time slots, these values are the integer times of value δ.

3.3. Appointment Scheduling Model

To deal with the appointment requests of CRs, we suggest a group of schemes for CPs, as shown in Figure 1. The goal of these schemes is to give the priority to the urgent CRs over the routine CRs for conducting the appointment scheduling. These schemes are prescribed in accordance with two key factors, *i.e.*, the CR type and the current number of idle CPs being available for appointments. The detailed description is as follows.

For the urgent CRs, they can exclusively obtain the appointment service of CPs even when they already have the scheduled appointment with the routine CRs. Specifically, if a new appointment is requested by a particular urgent CR and at least one CP is idle, this CP is allocated to the newly requested appointment. If there is no idle CP and at least one CP is allocated to a routine CR, this CP is re-allocated to the appointment newly requested by the particular urgent CR. Otherwise, the newly requested appointment is blocked. For the simplicity purposes, the scheduled appointment between the particular urgent CR and the allocated CP is called *ongoing urgent appointment*. For a new appointment requested by a particular routine CR, if at least one CP is idle, this CP is allocated to the newly requested appointment. Simply put, the scheduled appointment between the particular routine CR and the allocated CP is called *ongoing routine appointment*. Because the appointment requests of urgent CRs have the higher priority than the ones of routine CRs, a particular routine CR may be interrupted due to an appointment newly requested by an urgent CR. In this situation, the ongoing routine appointment associated with this particular CR needs to be paused.

The interrupted routine CR can request another idle CP with available time slots, the so-called *appointment handoff*. If the operation on appointment handoff is succeessful, the interrupted routine CR can resume the paused routine appointment. Otherwise, the paused routine appointment is forced to be terminated.

4. Queueing Modelling

Based on the above model of the appointment scheduling in eHealth, a DTMC queueing model is built up, as shown in **Figure 2**. This model has three characteristics: system state, state transition and steady-state probability [29] [30]. Let an integer pair (*i*, *j*) denote a system state when *i* ongoing routine appointments and *j* ongoing urgent appointments coexist in the system. The system state space is defined as $S = \{(i, j)\}$, where the values of *i* and *j* are constrained by $i \in [0, M]$, $j \in [0, M]$ and $(i + nj) \in [0, M]$.

4.1. State Transition

The state transition of the system is triggered by several activities. These are the activity of urgent CRs, activity of routine CRs without the appointment interruption by urgent CRs, the feedback of routine CRs in response to the appointment interruption.

4.1.1. Urgent Care Receiver Activity

The urgent CRs have two different activities:

- A new appointment *requested* by the urgent CR is assigned with a CP.
- The *urgent* CR releases the appointment from the CP at the service completion.

The first activity indicates the arrival of PU calls into the system. Both the second and the third activities indicate PU calls leaving the system.

For *j* ongoing urgent CRs in the system, we let λ_j^* and μ_j^*, denote the arrival and the leaving rates of PU calls, respectively. The values of λ_j^* and μ_j^* depends on the four parameters *j*, λ_1, μ_1 and *M*. They are given by:

$$\lambda_j^* = \begin{cases} \lambda_1, & 0 \le j \le M \\ 0, & \text{others} \end{cases} \quad (1)$$

$$\mu_j^* = \begin{cases} j\mu_1, & 0 \le j \le M \\ 0, & \text{others} \end{cases} \quad (2)$$

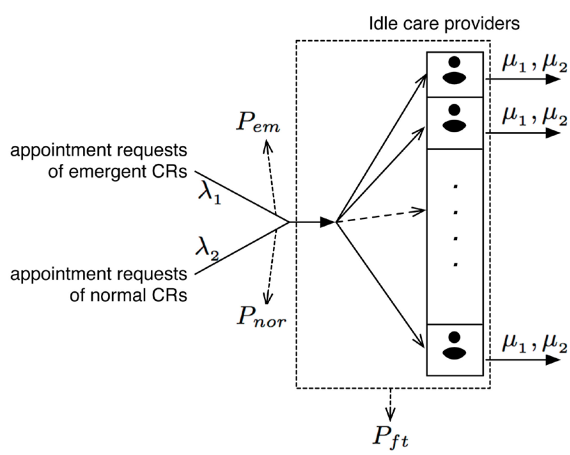

Figure 2. Queueing model with P_{ur} for blocking probability of urgent appointment requests, P_{ro} for blocking probability of routine appointment requests and P_{ft} for forced-termination probability of ongoing routine appointment.

4.1.2. Routine Care Receiver Activity without Interruption

Given that i ongoing routine CRs and j ongoing urgent CRs are in the system, a new appointment requested by a routine CR is treated in two different ways:

- For $(i + j) < M$, an idle CP is allocated to the routine CR with a scheduled appointment.
- For $(i + j) = M$, the appointment request is blocked.

Similar to the urgent CRs, we let $\lambda_{i,j}^{\dagger}$ denote the arrival rate of the appointment requests of the routine CRs. Let $\mu_{i,j}^{\dagger}$ denote the leaving rate of the routine CRs after completing the appointment. The values of $\lambda_{i,j}^{\dagger}$ and $\mu_{i,j}^{\dagger}$ depends on the six parameters i, j, λ_2, μ_2 and M. They are computed by:

$$\lambda_{i,j}^{\dagger} = \begin{cases} \lambda_2, & 0 \le i + j \le M \\ 0, & \text{others} \end{cases} \tag{3}$$

$$\mu_{i,j}^{\dagger} = \begin{cases} i\mu_2, & 0 \le i + j \le M \\ 0, & \text{others} \end{cases} \tag{4}$$

4.1.3. Routine Care Receiver Feedback in Response to the Appointment Interruption

Given the system at state (i, j), where $(i, j) \in S$ and $j < M$, a new appointment requested by an urgent CR is assigned with a particular CP. Because of the higher priority for urgent CRs, the routine CR being scheduled with this particular CP is interrupted and the associated ongoing routine appointment is paused. To resume the paused appointment, the interrupted routine CR attempts the handoff to another idle CP. The success in such an attempt depends on the numbers of ongoing urgent CRs and PU calls and connected SU calls in the system at state (i, j) as follows:

- If $(i + j) < M$, this means that the interrupted routine CR can find an idle CP for appointment scheduling. It is further assumed that a successful appointment handoff can be immediately accomplished by an interrupted routine CR. As a result, the system changes state from (i, j) to $(i, j + 1)$.
- If $(i + j) = M$, this indicates that the system becomes overloaded for re-allocating the CP to the interrupted routine CR. Therefore, the paused routine appointment is forced to be terminated. As a consequence, the system changes the state from (i, j) to $(i - 1, j + 1)$.

4.2. Steady-State Probability

The system state diagram is shown in Figure 3, where $\theta_{i,j}$ equals one if $\{j \neq m, n(m - 1) + q < (i + nj) \le (nm + q)\}$ and zero if others. $\varphi_{i,j}$ equals one if $\{i \neq 0, (i + nj) = (nm + q)\}$ and zero if others.

Let $\pi_{i,j}$ denote the steady-state probability of state (i, j). If $(i, j) \in S$, the value of $\pi_{i,j}$ is in the value range (0.0, 1.0). Otherwise, $\pi_{i,j}$ is equal to zero. Further, the rate of transition flow into a state (i, j) must be equal to the rate of transition flow out of this state.

For the four particular states (0, 0), $(M, 0)$ and $(0, M)$, we have Equations (5) - (7). For the states satisfying $0 < i < M$ and $j = 0$, we have the Equation (8).

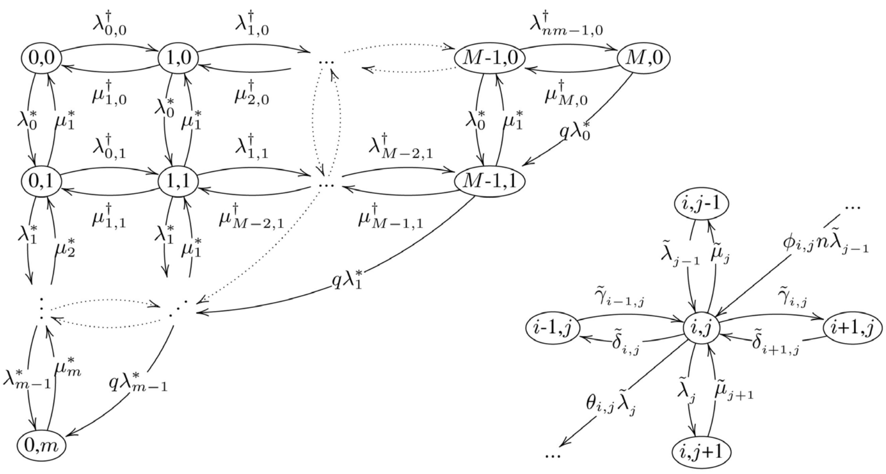

Figure 3. State diagram of the modeled system for $0 < g < m$, $0 < q < n$, $0 < h < q$.

For the states satisfying $i = 0$ and $0 < j < M$, we have the Equation (9). For the states satisfying $i + j = M$ and $0 < j < M$, we have the Equation (10). For other particular states, we have the Equation (11).

$$\pi_{0,0}\left(\lambda_{0,0}^{\dagger}+\lambda_0^{*}\right)=\pi_{1,0}\mu_{1,0}^{\dagger}+\pi_{0,1}\mu_1^{*} \tag{5}$$

$$\pi_{M,0}\left(\lambda_0^{*}+\mu_{M,0}^{\dagger}\right)=\pi_{M-1,0}\mu_{M-1,0}^{\dagger} \tag{6}$$

$$\pi_{0,M}\mu_M^{*}=\pi_{0,M-1}\lambda_{M-1}^{*}+\pi_{1,M-1}\lambda_{M-1}^{*} \tag{7}$$

$$\pi_{i,0}\left(\lambda_{i,0}^{\dagger}+\mu_{i,0}^{\dagger}+\lambda_0^{*}\right)=\pi_{i-1,0}\lambda_{i-1,0}^{\dagger}+\pi_{i+1,0}\mu_{i+1,0}^{\dagger},\ 0<i<M,\ j=0 \tag{8}$$

$$\pi_{0,j}\left(\lambda_{i,0}^{\dagger}+\lambda_j^{*}+\mu_j^{*}\right)=\pi_{1,j}\mu_{1,j}^{\dagger}+\pi_{0,j-1}\lambda_{j-1}^{*}+\pi_{0,j+1}\mu_{j+1}^{*},\ i=0,0<j<M \tag{9}$$

$$\pi_{i,j}\left(\lambda_j^{*}+\mu_j^{*}+\mu_{i,j}^{\dagger}\right)=\pi_{i-1,j}\lambda_{i-1,j}^{\dagger}+\pi_{i,j}\lambda_{j-1}^{*}+\pi_{i+1,j-1}\lambda_{j-1}^{*},\ i+j=M,\ 0<j<M \tag{10}$$

$$\pi_{i,j}\left(\lambda_{i,j}^{\dagger}+\mu_{i,j}^{\dagger}+\lambda_j^{*}+\mu_j^{*}\right)=\pi_{i-1,j}\lambda_{i-1,j}^{\dagger}+\pi_{i+1,j}\mu_{i+1,j}^{\dagger}+\pi_{i,j-1}\lambda_{j-1}^{\dagger}+\pi_{i,j+1}\mu_{j+1}^{*} \tag{11}$$

We sum up all steady-state probabilities in conjunction with $\sum_{\forall i,j}^{(i,j)\in S}\left[\pi_{i,j}\right]=1$. By combining the above equations, we can construct a set of linear equations. By solving them, we can accordingly compute the steady-state probabilities of all states.

4.3. Performance Metrics

The following performance metrics are considered.

4.3.1. Blocking Probability of Urgent Appointment

Clearly, the event of blocking the appointment requested by an urgent CR occurs for $j = M$. In other words, all the CPs are allocated to the urgent CRs and there is no ongoing routine appointment. Let P_{ur} denote the blocking probability of urgent appointment requests, and it is given by.

$$P_{ur}=\pi_{0,M} \tag{12}$$

4.3.2. Blocking Probability of Routine Appointment

According to the appointment handoff model, the event of blocking the appointment requested by a routine CR occurs for $(i + j) = M$. Let P_{ro} denote the blocking probability of routine appointment requests, and it is computed by.

$$P_{ro} = \sum_{\forall i,j}^{(i,j)\in S} \left[\pi_{i,j} \middle| (i+j) = M \right] \tag{13}$$

4.3.3. Forced-Termination Probability of Routine Appointment

For the state (i, j), when a particular CP is allocated to an urgent CR with the new appointment request, the termination of an ongoing routine appointment may occur for $i + j = M$ and $j < M$. This means the system has not enough CPs to accommodate the interrupted routine CR that has experienced the unsuccessful appointment handoff. Let P_{ft} denote the forced-termination probability of ongoing routine appointment. We then define P_{ft} as:

$$P_{ft} = \frac{\text{Total forced} - \text{termination rate of ongoing routine appointments}}{\text{Actual average appointment request rate of routine CRs}}$$

For the state $(i, j) \in S$ satisfying $i + j = M$, the forced-termination rate at this state is equal to the product of urgent appointment request rate and the number of terminated routine CRs. Given that the PU occupies the band with arrival rate λ_j^*, the number of terminated SU calls is equal to one. Therefore, at state (i, j) the forced termination rate is equal to λ_j^*. As a result, the total forced termination rate of ongoing calls equals to:

$$\sum_{\forall i,j}^{(i,j)\in S} \left[\lambda_j^* \pi_{i,j} \middle| (i+j) = M, j \neq M \right] \tag{14}$$

Because the routine appointment requests are blocked with probability P_{ro}, the actual average appointment request rate of routine CRs into the system equals $[\lambda_2(1 - P_{ro})]$. Subsequently, P_{ft} is given by:

$$P_{ft} = \frac{\sum_{\forall i,j}^{(i,j)\in S} \left[\lambda_j^* \pi_{i,j} \middle| (i+j) = M \right]}{\lambda_2 (1 - P_{ro})} \tag{15}$$

4.3.4. Appointment-Completion Throughput

For a particular category of CRs, the associated appointment-completion throughput is defined as the average rate of ongoing appointments completing with CPs. Let R_{em} and R_{ro} denote the appointment-completion throughput of urgent and routine CRs, respectively. According to the description in Section 3, each of appointments requested by different CRs faces four different cases (*i.e.*, being scheduled, being blocked, being force-terminated, and completing the service). Then, R_{em} and R_{ro} are computed by:

$$R_{ur} = \lambda_1 (1 - P_{ur}) \tag{16}$$

$$R_{ro} = \lambda_2 (1 - P_{ro})(1 - P_{ft}) \tag{17}$$

4.3.5. Average Service Time

The service time of a particular CR means the time spent by the allocated CP for dealing with the appointment scheduled with this particular CR. Let T_{ur} and T_{ro} denote the average service time of an ongoing appointment for the urgent and routine CR, respectively. To compute them, we need to consider the average numbers of ongoing urgent and routine CRs in the system, which are denoted by N_{em} and N_{ro}, respectively. According to the expectation definition, they are given by:

$$N_{ur} = \sum_{\forall i,j}^{(i,j)\in S} \left[i\pi_{i,j} \right] \tag{18}$$

$$N_{ro} = \sum_{\forall i,j}^{(i,j)\in S} \left[j\pi_{i,j} \right] \tag{19}$$

Subsequently, T_{ur} and T_{ro} can be computed with respect to Little's Theorem [30]:

$$T_{ur} = \frac{N_{ur}}{\lambda_1 \left(1 - P_{ur}\right)} \tag{20}$$

$$T_{ro} = \frac{N_{ro}}{\lambda_2 \left(1 - P_{ro}\right)\left(1 - P_{ft}\right)} \tag{21}$$

5. Performance Evaluation

This section reports on the performance evaluation of the modelled eHealth system for doing appointment scheduling.

5.1. Parameter Settings

To study the effects of the priority and handoff schemes on the appointment scheduling performance, both numerical analysis and simulation experiments are conducted.

Numerical analysis is carried out based on the developed queueing model. The corresponding parameter settings are reported in **Table 2**. Simulation experiments are carried out to demonstrate the validity of the numerical analysis. In our experiments, the simulator is developed in C/C++.

The results are shown in **Figures 4(a)-(f)**. In all figures, the marker "+" indicates the simulation result. From the tables and figures, we observe that the simulation results closely match the numerical results. The discussions of results are as follows.

Table 2. Parameter settings.

CPs	$M = 6$
Urgent CRs	$\lambda_1 \in \{4;6;8;10;12\}$ $\mu_1 = 5$
Routine CRs	$\lambda_2 = 6$ $\mu_2 = 6$

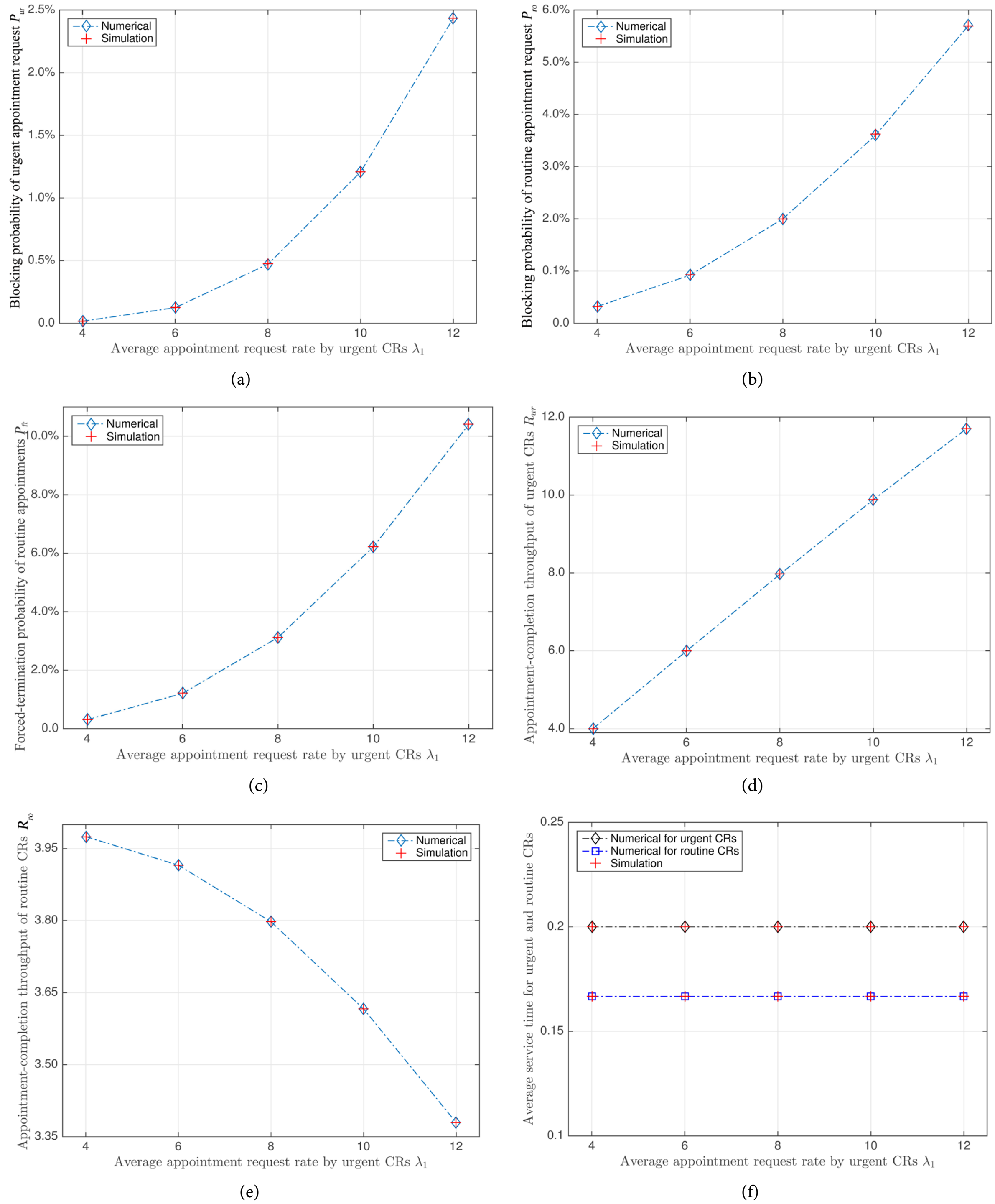

Figure 4. Numerical and simulation results of P_{ur}, P_{ro}, P_{ft}, R_{ur}, R_{ro}, T_{ur} and R_{ro}.

5.2. Blocking Probability of Appointment Requests

In Figure 4(a) and Figure 4(b), we observe that blocking probabilities of both urgent and routine appointment requests (*i.e.*, P_{ur} and P_{ro}) increase with the arrival rate of urgent appointment requests (*i.e.*, λ_1). The reasons for this under two different types of CRs are different.

For the urgent CRs, they can exclusively occupy the idle time slots provided by CRs, which are the limited resource with a fixed max value, *i.e.*, $M = 6$. Therefore, the more the appointments are requested by urgent CRs, the higher the possibility of experiencing the blocking event becomes. For the routine CRs, they are only allowed to opportunistically share the CRs for appointment scheduling, when these CRs are not allocated to the urgent CRs. In this situation, the system needs to deal with more urgent appointments. As a consequence, the resource availability for routine CRs is decreased with λ_1, and thus a larger amount of routine appointment requests are accordingly blocked.

5.3. Forced-Termination Probability of Routine Appointments

As addressed in Subsection 3.3, the forced-termination of an ongoing routine appointment occurs for the case of no available CPs for accomplishing the appointment handoff. As shown in Figure 4(c), the forced-termination probability (*i.e.*, P_{ft}) increases with λ_1. This is because the more urgent CRs are requesting for resources (*i.e.*, CPs), and thus leading to the more interruptions caused to the ongoing routine appointments. As such, the system capability of dealing with the appointment handoff for the interrupted routine appointments is decreased.

5.4. Appointment-Completion Throughput

Figure 4(d) and Figure 4(e) show that the appointment-completion throughs (*i.e.*, R_{ur} and R_{ro}) increases and decreases with λ_1 for the urgent and routine CRs, respectively. This is because when λ_1 increases, a larger amount of urgent appointments can be actually initiated in the system. As a result, more urgent appointments can be accomplished. On the contrary, the amount of initiated routine appointments becomes smaller, while fewer routine appointments can be accomplished.

5.5. Average Service Time

Figure 4(f) shows that the average service times (*i.e.*, R_{ur} and R_{ro}) maintain as constant values for the urgent and routine CRs. The reason for this result is intuitive. For the urgent CRs, they can request the appointments from CPs by ignoring the existence of the routine appointments. Therefore, the corresponding scheduling service provided by the system can be modelled as M/M/m-lose system, where the average service time only depends on the service rate [29]. Similarly, although the scheduling service for the routine CRs is constrained under the exclusive occupancy by the urgent CRs, the corresponding service model can also be modelled as M/M/m-lose system.

6. Conclusions

The appointment scheduling in the healthcare system was studied in our paper. Specifically, the heterogeneous aspect of Care Receivers (CR)s was considered in terms of urgent and routine users. The scenario of the concurrent appointment scheduling on these two different types of CRs was presented. The importance of giving the higher priority to the urgent CRs over the routine CRs was addressed as well. To provide such priority, the interruption activity on the ongoing routine appointments due to the newly arrived urgent CRs was discussed. The appointment handoff mechanism was therefore suggested for the interrupted routine appointments.

To investigate the scheduling performance in the considered system, we used a Discrete Time Markov Chains (DTMC) based queueing model. We presented a numerical solution to this model. For performance evaluation, we derived seven different metrics. They are the blocking probabilities of urgent and routine appointments requests, the forced-termination probability of ongoing routine appointments, the appointment completion throughputs of urgent and routine CRs, the average service time of urgent and routine appointments. The numerical results were also validated by the simulation experiments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Active, A.G.E. (2011) Prague Local Action Plan: Age and Care. Urbact II Thematic Network.

[2] Michael, J., Goran, M.D. and Jean, S. (2001) E-Health: Restructuring Care Delivery in the Internet Age. *Journal of Healthcare Information Management*, **15**, 3-12.

[3] European Commission (2012) eHealth Action Plan 2012-2020: Innovative Healthcare for the 21st Century.

[4] Stroetmann, K.A., Jones, T., Dobrev, A. and Stroetmann, V.N. (2006) eHealth Is Worth: The Economic Benefits of Implemented eHealth Solutions at ten European Sites. Technical Report, European Commission.

[5] Guo, Y., Yao, Y. and Bai, G.H. (2016) On Enhancement of Inter-Activity for Knowledge Sharing in eHealth. 2016 *International Conference on Communications*, Bucharest, Romania, 9-10 June 2016, 247-250. https://doi.org/10.1109/ICComm.2016.7528335

[6] Gupta, D. and Denton, B. (2008) Appointment Scheduling in Health Care: Challenges and Opportunities. *IIT Transactions*, **40**, 800-819. https://doi.org/10.1080/07408170802165880

[7] Feldman, J., Liu, N., Topaloglu, H. and Ziya, S. (2014) Appointment Scheduling under Patient Preference and No-Show Behavior. *Operations Research*, **62**, 713-972. https://doi.org/10.1287/opre.2014.1286

[8] Haji, M. and Darabi, H. (2011) A Simulation Case Study: Reducing Outpatient Waiting Time of Otolaryngology Care Services Using VBA. 2011 *IEEE International Conference on Automation Science and Engineering*, Trieste, Italy, 24-27 August 2011, 525-530. https://doi.org/10.1109/CASE.2011.6042448

[9] Song, M. and Zhang, Y. (2014) Queueing Approach for Social Welfare under Appointment Delays. 2014 11*th International Conference on Service Systems and Service Management*, Beijing, China, June, 25-27 June 2014, 1-5.

[10] Poon, C.C.Y., Liu, Q., Gao, H., Lin, W.H. and Zhang, Y.T. (2011) Wearable Intelligent Systems for E-Health. *Journal of Computing Science and Engineering*, **5**, 246-256. https://doi.org/10.5626/JCSE.2011.5.3.246

[11] Lian, J., Distefano, K., Shields, S.D., Heinichen, C. and Giampietri, M. (2010) Clinical Appointment Process: Improvement through Schedule Defragmentation. *IEEE Engineering in Medicine and Biology Magazine*, **29**, 127-134. https://doi.org/10.1109/MEMB.2009.935718

[12] Premarathne, U.S., Han, F., Khalil, I. and Tari, Z. (2013) Preference Based Load Balancing as An Outpatient Appointment Scheduling Aid. 35*th Annual International Conference of the IEEE EMBS*, Osaka, Japan, 3-7 July 2013, 1298-1301. https://doi.org/10.1109/EMBC.2013.6609746

[13] Le, M.-D., Nguyen, M.H.N., Baril, C., Gascon, V. and Dinh, T.B. (2015) Heuristics to Solve Appointment Scheduling In Chemotherapy. *IEEE International Conference on Computing & Communication Technologies Research*, *Innovation*, *and Vision for Future*, Can Tho, Vietnam, 25-28 January 2015, 59-64.

[14] Mardiah, F.P. and Basri, M.H. (2013) The Analysis of Appointment System to Reduce Outpatient Waiting Time at Indonesia's Public Hospital. *Human Resource Management Research*, **3**, 27-33.

[15] Alfonso, E., Xie, X. and Augusto, V. (2015) A Simulation-Optimization Approach for Capacity Planning and Appointment Scheduling of Blood Donors Based on Mathematical Programming Representation of Event Dynamic. 2015 *IEEE International Conference on Automation Science and Engineering*, Gothenburg, Sweden, 24-28 August 2015, 728-733. https://doi.org/10.1109/CoASE.2015.7294167

[16] Kim, S. and Giachetti, R.E. (2006) A Stochastic Mathematical Appointment Overbooking Model for Healthcare Providers to Improve Profits. *IEEE Transactions on Systems*, *Man*, *and Cybernetics Part A*: *Systems and Humans*, **36**, 1211-1219. https://doi.org/10.1109/TSMCA.2006.878970

[17] Rohleder, T. and Klassen, K. (2000) Using Client-Variance Information to Improve Dynamic Appointment Scheduling Performance. *Omega*, **28**, 293-302. https://doi.org/10.1016/S0305-0483(99)00040-7

[18] Wang, W. and Gupta, D. (2011) Adaptive Appointment Systems with Patient Preferences. *Manufacturing & Service Operations Management*, **13**, 281-418. https://doi.org/10.1287/msom.1110.0332

[19] Gupta, D. and Wang, W. (2011) Patient Appointments in Ambulatory Care. In: Hall, R., Ed., *Handbook of Healthcare System Scheduling*: *Delivering Care When and Where It Is Needed*, Springer, New York, 65-104. https://doi.org/10.1007/978-1-4614-1734-7_4

[20] Wijewickrama, A.K.A. and Takakuwa, S. (2006) Simulation Analysis of an Outpatient Department of Internal Medicine in a University Hospital. *Proceedings of the* 2006 *Winter Simulation Conference*, Monterey, CA, 3-6 December 2006, 425-432. https://doi.org/10.1109/WSC.2006.323111

[21] Zhang, Z., Xie, X. and Geng, N. (2014) Simulation-Based Surgery Appointment Sequencing and Scheduling of Multiple Operating Rooms. 2014 *IEEE International Conference on Automation Science and Engineering*, Taipei, Taiwan, 18-22 August 2014, 399-404. https://doi.org/10.1109/CoASE.2014.6899356

[22] Cola, C. and Valean, H. (2015) E-Health Appointment Solution, A Web Based Approach. 2015 *E-Health and Bioengineering Conference*, Romania, 19-21 November 2015, 1-4. https://doi.org/10.1109/EHB.2015.7391431

[23] Sankarananrayanan, S. and Wani, S.M.A. (2014) NFC Enabled Intelligent Hospital

Appointment and Medication Scheduling. 2014 2*nd International Conference on Information and Communication Technology*, Bandung, Indonesia, 28-30 May 2014, 24-29. https://doi.org/10.1109/ICoICT.2014.6914034

[24] Mey, Y.S. and Sankaranarayanan, S. (2013) Near Field Communication based Patient Appointment. 2013 *International Conference on Cloud & Ubiquitous Computing & Emerging Technologies*, Pune, India, 15-16 November 2013, 98-103. https://doi.org/10.1109/CUBE.2013.27

[25] Liu, Y., Geng, N. and Zhu, Y. (2015) MDP-Based Outpatient Scheduling for Multiple Examinations. *IEEE International Conference on Industrial Engineering and Engineering Management*, Singapore, 6-9 December 2015, 1312-1317. https://doi.org/10.1109/IEEM.2015.7385860

[26] Walker, D., Shanks, E., Montoya, D., Weiman, C., Pérez, E. and De Pagter, L. (2015) Towards a Simulation-Based Methodology for Scheduling Patient and Providers at Outpatient Clinics. *Winter Simulation Conference*, Huntington Beach, CA, 6-9 December 2015, 1515-1524. https://doi.org/10.1109/WSC.2015.7408273

[27] Sehr, M.A., Bitmead, R.R. and Fontanesi, J. (2015) Multi-Class Appointments in Individualized Healthcare: Analysis for Scheduling Rules. 2015 *European Control Conference*, Linz, 15-17 July 2015, 1225-1230. https://doi.org/10.1109/ECC.2015.7330707

[28] Yan, C., Tang, J. and Zhu, H. (2013) Appointment Scheduling for Routine and Urgent Patients under Deterministic Service Time. 2013 25*th Chinese Control and Decision Conference*, Guiyang, 25-27 May 2013, 1270-1274. https://doi.org/10.1109/CCDC.2013.6561120

[29] Kleinrock, L. (1975) Part I: Elementary Queueing Theory. In: *Queueing Systems*, John Wiley & Sons, Inc., Hoboken, NJ, 89-147.

[30] Trivedi, K.S. (2002) Probability and Statistics with Reliability, Queuing, and Computer Science Applications. John Wiley & Sons, Hoboken, NJ.