Simulation Optimization of Blocking Appointment Scheduling Policies for Multi-Clinic Appointments in Centralized Scheduling Systems

Nimish Dharmadhikari, Dr. Jun Zhang
North Dakota State University, Fargo, ND 58108-6050

Abstract—The development of methodology to schedule multi-clinic appointments for patients in a complex hospital setting is important. This paper develops Block Scheduling with Priority (BSP) policy for multi-clinic appointment scheduling in the centralized scheduling system (CSS). A heuristic approach involving simulation optimization is proposed. The average reward per patient performance measure is used to evaluate the effectiveness of the proposed scheduling policies. The proposed policy is compared with the frequently used first come first served (FCFS) scheduling policy. The results from simulation experiments suggest that the BSP policy ensure superior performance than FCFS policy when the demand of multi-clinic appointments is higher.

Index Terms—Simulation optimization, centralized scheduling, multi-clinic appointments, block scheduling with Priority

I. INTRODUCTION

Most complex hospitals in developed countries require patients to make appointments in advance to gain access for healthcare service. Therefore, improvement of patient appointment scheduling system is essential to obtain efficient and effective operation of complex healthcare systems.

Literature review [1] [2] has shown that centralized scheduling system has gained advantages for complex hospital systems that consist of multiple interrelated clinics. However, scheduling policies for centralized scheduling system are seldom studied in literature. Therefore, the paper proposes a simulation-based scheduling policy that can benefit the hospitals with patients required multiple appointments in same day visit.

The solution can be widely used in various hospital settings. For instance, in many Veterans Affairs (VA) hospitals, many patients have to travel long distances to the nearest VA hospitals. And most of those patients are seniors and have multiple health conditions. They have to see different providers for their illness. For examples, they need to see their primary care doctors and dentists. Due to the long travel distance, patients with multiple-appointment needs are willing to visit the multiple clinics in the same day. If they don’t get the appointments in the same day, they might make late cancellation or do not show up. This will cause providers’ idle time due to the cancellation or no-show, and rescheduling, which will increase the hospitals’ operation cost. The first come first served (FCFS) scheduling policy used in the hospitals cannot provide preferred timeslots allocated to patients with multi-clinic appointments. The similar problem can be found in many hospital systems worldwide. Therefore, this paper proposes block scheduling with priority (BSP) heuristics scheduling policy to address this problem.

Patient appointment scheduling controls the overall efficiency and appropriate access to the patient healthcare [3]. Overall efficiency of hospital system depends upon the proper utilization of the healthcare resources such as providers, nurses and equipment. Timely access is important for the outcome of the treatment as well as the patient satisfaction. This is provided by the correct patient appointment scheduling. This literature review is focused on the study of the past efforts concentrated on the patient appointment scheduling.

Patient appointment scheduling has been a topic of interest since the foundational work of Bailey [4] which has focused on the procedure for the reduction of the patient waiting time in the system. Welch further improves this concept with their conclusion on the scheduling of the patients in the block [5]. A fixed time interval considered for the appointment scheduling is called the block. The appointment scheduling policies are summarized in the Table 1. They are classified into three groups: 1) Number of patients present in the block, 2) Patient arrival rules, and 3) Scheduling rules.

Block presented in the available literature is used only for the single clinic scheduling. Progressive block was used by Snow et al. [16]. They kept 10% timeslots block for new patients every month. This was performed till they reach 50% open timeslots in a week for new patients. This method was successful for open access system in their clinic. No efforts found giving preference to the patient needing same day appointments in multiple clinics according to his/her need.

<table>
<thead>
<tr>
<th>Policies</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients present in the block:</td>
<td>Bailey [4]; Welch [5]; Ho and Lau [6]; Cayirli and Veral [7]; Vermeulen, Bohte, Somefun, and Poutre [8]; Gupta and Denton [1]; Vermeulen, et al. [9] and Wijewickrama and Takakuwa [10]</td>
</tr>
<tr>
<td>All patients in single block</td>
<td></td>
</tr>
<tr>
<td>All patients in individual blocks</td>
<td></td>
</tr>
<tr>
<td>Multiple patients in variable blocks</td>
<td></td>
</tr>
</tbody>
</table>
2. Patient arrival rules:
   * All patients arrive at the starting of the block
   * Patients arrive randomly throughout the block
   * Patients arrive at constant interval (their appointment time)

   - Bailey [4]; Welch [5]; Ho and Lau [6]; Cayirli and Veral [7]; Kolisch and Sickingr [11]; Klassen and Rohleder [12]; Cayirli, Veral, and Rosen, [13]; Wijewickrama and Takakuwa, [10]; Vermeulen, Bothe, Somefun, and La Poutré [8]; and Vermeulen, et al. [9]

3. Scheduling rule:
   * First Come First Server/First In First Out
   * First come random served
   * Other specialized systems

   - Paulussen, Jennings, Decker, and Heinzl [14]; Hong and Prabhu [15]; Paulussen et al. [16]; Klassen and Rohleder [12]; Cayirli, Veral, and Rosen [13]; Vermeulen, Bothe, Somefun, and La Poutré [8]; Jula and Leachman [17]; Kolisch and Sickingr [11]; Vermeulen, et al. [9]; and Wijewickrama and Takakuwa [10]

This paper is organized as follows: Section II describes the scheduling problem. Section III proposes the heuristic block scheduling with priority (BSP) policy. Simulation experimentation is discussed in the section IV. Section V concludes the paper with future development possibilities.

II. PROBLEM STATEMENT

This paper studies a complex hospital system consisting of n clinics such as primary care clinic, audiology clinic, neurology clinic, or cardiology clinic, etc. The hospital uses the centralized patient appointment scheduling system as it is best suited for the current complex hospital system. A simple centralized scheduling system is depicted in Fig. 1. Scheduling clerks in the centralized scheduling department receive scheduling requests from patients. They will match patients’ appointments requests with the providers’ available timeslots.

Appointment requests of patients arrive randomly, but the interarrival times follow a known probability distribution. Each patient might need appointment in one clinic, two clinics, or n clinics according to the need. Patients with multiple appointments prefer their appointments in the same day. If they do not get appointments in the same day, they might make late cancellation or do not show up. This will cause idle time of resources and reschedule so as to increase the operational cost of the hospital. Therefore, a cost will occur if patients with multiple appointments cannot obtain all the required appointments in the same day.

Currently, FCFS scheduling policy is used to assign timeslots to appointment requests in each clinic. However, this policy cannot provide preferred appointments to patients needing appointments in multiple clinics in single day. Therefore, this paper proposes a heuristics block scheduling policy in order to increase the possibility that patients with multi-clinic appointments can obtain same day appointment. This policy will reduce the total operation cost of the hospital by reducing the possibility of late cancellation or no-show.

The following section will discuss the proposed scheduling policy in detail.

III. PROPOSED BLOCK APPOINTMENT SCHEDULING POLICY

This paper proposes an appointment scheduling policy block scheduling with priority (BSP). A specific time allocated for the patient’s visit to provider is called appointment timeslot. The block is the specific appointment timeslot or number of timeslots reserved for the patients with preference. For example, in Fig. 2 two different provider calendars are shown. One calendar is for clinic 1 which has first four timeslots open for any patient and last three are reserved as block. Second calendar has first three timeslots are reserved for patients with priority. This is example of block.

Patients’ preferences about the appointment provide priority for scheduling. This priority is about appointments needed in number of clinics and the week of appointment. Number of clinics range from one clinic, two clinics, or multiple clinics. The number of weeks range from week 1 to week z. If a patient wants a multi-clinic appointment he/she gets the priority over the patient with the single clinic appointment. Similarly a patient has higher priority for week 1 appointment over appointments in later weeks. Patients requesting multi-clinic appointment with priority are scheduled in the block. Different timeslots are reserved as block for different clinics so they do not overlap. For example in the Fig. 2, clinic 1 has open timeslots in the start of the day while clinic 2 has open timeslots at the end of the day to avoid overlap.
Number of reserved timeslots as block is the most important parameter while considering BSP. If the block has more timeslots reserved than required, it will increase the resources idle time and decrease efficiency. Similarly if the number of timeslots reserved is less than required, patients will not get the same day appointment. Thus it is very important to develop the correct block structure.

Steps for developing a block structure are as follows.

Step 1: Determine number of clinics in the hospital.

Step 2: Determine the interactions between the clinics. Interactions mean the sharing of patients between the clinics. If a high number of patients needs the appointments in any two clinics on same day; this is considered as a high interaction between those two clinics. Similar interactions are found in all major clinics.

Step 3: Reserve timeslots as block in different clinics for multi-clinic appointments avoiding overlaps. Various block structures are proposed for different scenarios in the clinic. For example various percentages of total timeslots can be reserved as block like 10%, 20%, etc. Also it can be performed with different number of timeslots reserved as block such as 1, 2, 3, or more timeslots reserved as block each day.

Step 4: Perform the simulation to test different strategies proposed in step 3. This gives the best suitable block strategy for the particular scenario. Various scenarios tested are combination of different interarrival times of patients, different demand cases for multi-clinic appointments, and different block structures. It helps to make the rest of scheduling decisions.

Step 5: Using the best strategy found out with simulation study complete scheduling procedure is defined for the multi-clinic hospital scheduling. Block scheduling starts with the patient’s appointment request entering in the system. Patient requests the appointment from the scheduling window of ‘z’ weeks. This is performed with the probabilities for different weeks as probability (Pi): week number (Vi). This stage helps to prioritize the patients. The policy specifies that if the appointment is in first week, it is higher priority and if it is in the zth week this priority is least. This can be assumed to be an emergency visit for the patient. If the patient needs an appointment in the earlier week, it is assumed that he/she needs it urgently. Then patient’s choice for single clinic or two clinics or n clinics appointment is considered. This is also used to prioritize the patients. If the patient needs a single clinic appointment, it is not scheduled in block. If he/she needs appointment in more than one clinic then it is scheduled in block. At the end of the scheduling process, the performance of the system is measured with the help of average reward performance measure.

A. Performance Measure

A monetary term of total reward is developed for this study. Total reward consists of revenue and costs associated with the patient visit. If revenue is considered alone in the research, similar results for all the situations are expected even if patients need different types of appointments. In order to maintain the diverse structure of appointments and to keep convergence with the real life scenarios, different costs associated with the scheduled and the unscheduled appointments are added in the model. These costs are subtracted from the revenue generated to get the total reward. Total reward can be written as the subtraction of various costs like patient multi-clinic appointment cost, provider idle cost, provider overtime cost, and patient waiting cost from the total revenue.

\[ T = R - (M + W + I) \]

(1)

Where

\( T = \text{Total Reward} \)
\( R = \text{Revenue Generated From Patient Visit} \)
\( M = \text{Mismatch Appointment Cost} \)
\( W = \text{Patient Waiting Cost} \)
\( I = \text{Provider Idle Cost} \)

To evaluate the performance of the model in the current operating conditions a performance measure the average reward per appointment (A) is developed. This performance measure is calculated by dividing the total reward at the end of model run (T) with the total number of appointments booked at the end (N). This figure is calculated for all models and with different set of runs thus the means are generated for all the runs. These all means are plotted and compared and the relative performances of all the three models are calculated.

\[ A = \frac{T}{N} \]

(2)

Where

\( A = \text{Average Reward per Appointment} \)
\( T = \text{Total Reward} \)
\( N = \text{Number of Booked Appointments} \)

IV. EXPERIMENTAL ANALYSIS

A case study of a hospital with the higher level of interactions between three clinics is presented in this paper. Patients tend to schedule appointments in these clinics on the same day. Considering the diverse need for care, various types of patients are present in the hospital system at the same time, for example return patients, emergency patients, or new patients [19]. This case study focuses only on return patients as the largest number of appointments scheduled are of this type. The scheduling system can be depicted in the Fig. 3.

A Patient calls the centralized scheduling department. The scheduler checks the available timeslots for clinic 1, 2 and 3 and tries to book the appointment. Every provider has a 30 minutes timeslot for each patient. This appointment time is fixed. As only return patients are considered in this research, there is no need to differentiate in the time allocated to the patients.

The major part in the appointment scheduling for patients is matching demand with the supply of available timeslots for appointments. The supply depends on the providers’ schedule. All providers are considered to work five days in each week for same time and with same break in between.
Providers work from Monday to Friday in an 8-hour shift with a one-hour lunch break. Every day, the working shift starts at 8:00 and ends at 16:00 with a 11:30 to 12:30 lunch break. Nurses are assumed to finish their work before provider start. With 30 minutes timeslots, as stated earlier, the scheduler can schedule 14 appointment timeslots for each provider, each day of clinic operation.

Fig. 3 Scheduling System

Patients request for the scheduling of the clinic appointments by telephone or printed request from provider. This request arrival process is assumed as Poisson process [20]. The interarrival time between each request is distributed with exponential ‘x’ minutes. While performing the simulation experiments, the value of λ is varied to get results for different values. Duration of each simulation run is to be set as one year. This run will consist of 52 weeks. Scheduling window is measured as three months that is 12 weeks. This scheduling window is dynamic as week 1 is over, scheduling for week 13th is started with week 2 is starting week and 13 as end week.

This research conducts appointment scheduling taking into account patient’s demand pattern for the appointment(s). A patient has first choice of one of the three types of appointments: 1) Single clinic appointment which will consist of choice of appointment for either clinic one, two or three 2) two clinics appointments, in which patient will choose appointments for either clinic one and two or clinic one and three or clinic two and three 3) three clinics appointment in which a patient will request to schedule appointment for all three clinics in single day. After that the patient can prefer a particular week from the 12 week scheduling window. This process is performed with certain probabilities. As stated earlier these probabilities are given in format as probability (Pi): week number (Vi). For example: (P1:V1, P2:V2 ... P12:V12).

There are four different demand cases for simulation experiments which are explained as follows. Demand case 1: Low demand for multi-clinic appointment. In this case there is high demand for the single clinic appointment which is 80% of total demand and low demand which is 10% for each two and three clinics appointments. Demand case 2: Moderate demand for appointment. This case states that there is moderate demand for all types of appointments. Therefore it has around 33.33% demand for single clinic, two clinic, or three clinic appointments. Demand case 3: High demand for two clinic appointment. Demand for two clinic appointments is 80% of total demand. Demand case 4: High demand for three clinic appointment. Demand for three clinic appointment is 80% of total demand.

These demand cases are further distributed in different structures of block strategy which are: FCFS, block 2 timeslots, block 3 timeslots, block 4 timeslots, block 5 timeslots, block 7 timeslots, and block 8 timeslots. These different scenarios provide enough variations in the system and block structure. Each combination was run for 10 replications. Total number of runs for the complete experimental set up was: number of strategies 7 * cases 4 * times 4 * replications 10 = 1120 runs.

Graphs of average reward plotted against different interarrival time are shown in Fig. 4 to Fig. 7. These graphs compare average reward for number of blocking structures. They show changes in the average reward as the blocking changes and interarrival time between patients change.

Statistical tests are used to compare two sample means and examine if they are different from each other. The pairwise comparison is performed as a t-test with sample of size 10 each. The t-test results show that the sample means are different than each other or they are same. Null hypothesis is stated as the means are same. α is considered as the 0.05 so if the p-value is greater than 0.05 then we cannot reject the null hypothesis. Some sample results are presented in Tables 2 and 3.

Table 2 Statistical test for Block 2 timeslots in Demand Case 2

<table>
<thead>
<tr>
<th>Interarrival time 20 minutes</th>
<th>NSP</th>
<th>FCFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>106.7</td>
<td>108.1</td>
</tr>
<tr>
<td>Variance</td>
<td>2.45555</td>
<td>3.43333</td>
</tr>
<tr>
<td>Observations</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.35588</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>-2.2644</td>
<td></td>
</tr>
<tr>
<td>P(&gt;</td>
<td>=t</td>
<td>one-tail)</td>
</tr>
</tbody>
</table>

Table 3 Statistical test for Block 5 and 7 timeslots in Demand Case 2

<table>
<thead>
<tr>
<th>Interarrival time 15 minutes</th>
<th>Block 5</th>
<th>Block 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>130.1</td>
<td>130.3</td>
</tr>
<tr>
<td>Variance</td>
<td>3.877778</td>
<td>4.9</td>
</tr>
<tr>
<td>Observations</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.70607</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
These results are summarized in the following discussion:

- In demand cases 2, 3, and 4, the BSP performs better than the FCFS. This is shown in Fig. 5, 6, and 7.
- Fig. 4 depicts if there is lower demand for multi-clinic appointments (demand case 1), the FCFS outperforms the BSP. This is expected as BSP is developed for the major function of scheduling multi-clinic appointments.
- Graphs show that in most cases there is no major difference between block 5 timeslots and block 7 timeslots. This can be explained as the revenues generated by those strategies are comparable and cost structures in most cases are similar.
- Pairwise comparison tests were performed. Samples of these tests can be seen in Tables 2 and 3. The results from these tests showed that the respective p-values are less than 0.05. This proposes that the sample means are different and thus the two samples are different from each other.
- Blocking strategy outperforms FCFS strategy in the higher demand for multi-clinic appointments. Graphs for demand cases 3 and 4 fortify these results.

V. CONCLUSION AND FUTURE DEVELOPMENT

This paper develops heuristic block scheduling with priority (BSP) policy. This policy is developed for the scheduling in complex hospital settings. In complex hospitals patients commonly require more than one appointment in different clinics on the same day. The traditional first come first served (FCFS) policy often cannot provide same day appointments for patients. BSP policy attempts to develop the block structure for patients requesting the multi-clinic appointments. This policy helps to schedule multi-clinic appointments in single day.

The simulation models were developed for comparison between FCFS and BSP. These models were run for different experimental scenarios. A performance measure of average reward is developed, considering the monetary reward for a patient visit and different costs such as idle, waiting, and mismatch costs. This performance measure was used to test...
the performance of the models developed with the simulation. The experiments were carried out to test and compare the performance of the two models in different demand cases: patient arrival rates and different blocking strategies. The results from these experiments were in accordance with the expectations. In case of lower demand for multi-clinic appointments, FCFS is the optimal strategy. BSP is the optimal strategy in case of high demand for multi-clinic appointments.

This research is based on the assumption that the patients have similar appointment time requirements. A study involving diverse patients with different appointment time requirements could be performed. For example, new patients with one hour appointments or non-critical patients with 15 minute appointments could be used in the simulation. Such a study would test the usefulness of the blocking strategy in different scenarios. This research involves simple discrete event simulation techniques. This simulation could be expanded with the help of agent based modeling. Intelligent agents are important tool to capture the interactions between complex situations. The agent based modeling will help to integrate the current clinic system. This would expand the model and provide improved examination.

REFERENCES


AUTHOR BIOGRAPHY

Nimish Dharmadhikari is Ph.D. student in Transportation and Logistics program at North Dakota State University. He has completed his MS in Industrial Engineering and Management from North Dakota State University. His research interests are optimization, simulation, Geographical Information System (GIS), supply chain analysis, economic modeling, and agricultural transportation.

Dr. Jun Zhang is Assistant Professor in Department of Industrial and Manufacturing Engineering at North Dakota State University. She completed her Ph.D. in Industrial Engineering from Purdue University. Her research interests are Lean manufacturing and logistics, Production planning and inventory control, Scheduling, Simulation optimization, Models and methodologies of stochastic optimization, Healthcare engineering, Facility design, Supply chain management, Artificial intelligence, Machine learning and data mining, Computer integrated manufacturing.