EVALUATION OF VARIOUS SCHEDULING DISCIPLINES BY COMPUTER SYSTEMS

Benjamin Lev and Robert J. Caltagirone

Temple University

ABSTRACT

The assignment of patients to examination rooms in a diagnostic radiology department can be categorized as a job-shop scheduling problem. This paper shows a way of using a GPSS (General Purpose Simulation System) discrete event simulation model to test, analyze and find the best heuristic scheduling algorithms that can be applied to patient flow in a diagnostic radiology department. Four main measures of performance are considered; waiting time before examination, total time in the system, utilization of staff and equipment, and the number of patients in the system at the end of a working day. Among the scheduling disciplines tested include the following six; smallest number of patients in each queue, smallest work load in each queue, shortest processing time (SPT), truncated SPT (with priorities), common queue (patients join a single queue versus multiple queues based on examination type), and truncated common queue. Numerical data of Temple University's Radiology Department are included, and results indicate that two scheduling disciplines are superior to the others. Finally, a feasibility study of implementing the findings with and without a computer are discussed.

I. INTRODUCTION

The operation of a radiology department can be viewed as a complex queueing network where each patient passes through a series of service stations in order to complete his processing. Analytical models have been developed for some multiple-server queueing systems, however, several characteristics of the radiology department make it difficult, if not impossible, to apply such models. Each server in the department is not identical. For example, some examination rooms are dedicated solely to one type of radiological study, while others can perform many types. Patient flow in the department is subject to priorities that can be changed depending on system status. In addition, the structure of the system does not conform to the standard M/M/S models developed in the literature. Input in such systems is served by several parallel service stations, each of which performs the identical service. The radiology department consists of several different service stations, some of which are arranged in tandem, while others are parallel. To analyze such a system requires the development of a computer simulation model that can take into account the combined effects of the interactions between patients and system entities.

A discrete event simulation model of patient flow in a radiology department was written using GPSS simulation language with FORTRAN subroutines incorporated to handle complex decision processes and data manipulation. The program consists of approximately 750 GPSS blocks and 300 FORTRAN statements. It is run on an IBM 370/165 computer requiring 300K bytes of core. Each run represents one working day (160 patients) and costs about 10 dollars. The model was used to evaluate several patient scheduling procedures which assign patients to examination rooms given that the patient has arrived to the department. Each scheduling procedure was evaluated in terms of parameters which describe the performance of the system. Four main measures were considered:

1. Waiting time prior to examination.
2. Total time in the system.
3. The distributions of waiting time and total time.
4. The number of patients in the system at 4:30 PM.

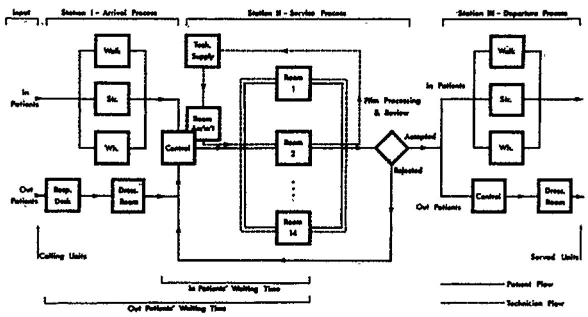
Based on these measures a "best" procedure was determined.

II. DESCRIPTION OF SYSTEM

The radiology department under study is that of Temple University's Hospital. It serves approximately 160 patients per day and consists of 14 examination rooms, 10 x-ray technicians, 10 orderlies and 16 dressing rooms. (This represents the core radiology department not including special studies, emergency, portables, etc.). In addition, several personnel are involved in administering patients as they arrive to the department. Illustration 1 represents the flow of patients through the department.

ILLUSTRATION 1

Patient Flow



Patients are divided into inpatients (65%) and outpatients (35%). Outpatients report to the reception desk where initial data is collected. Some of the outpatients have to dress for the examination (80%), while others report directly to the control desk. At the control desk a room assignment is made based on patient characteristics and the status of the system. Various assignment or scheduling policies can be evaluated to see which produces the best result in terms of system performance. Once an assignment is made the patient waits until the room and necessary personnel become available. When the room and personnel are available the examination is conducted and the patient returns to the waiting room until his film is processed and checked for technical quality. If the film is acceptable the patient reports to the reception desk for final processing and departs from the system. Approximately 5% of the films are rejected. These patients re-enter the system at the control desk with a higher priority assigned and repeat the processing.

Inpatients are escorted to and from the department by orderlies. The mode of transportation can ambulatory (5%), wheelchair (66%), or stretcher (29%). Upon arrival inpatients report to the control desk where the room assignment is made. From this point until departure processing is the same as that of outpatients. Repeats are recycled while completed patients are escorted back to their room.

III. DESIGN OF EXPERIMENT

Details of the simulation model itself have been reported on elsewhere (2). The objective of this analysis is to improve the operating efficiency of the radiology department. Efficiency here is difficult to define. It might be any combination of the following factors:

1. The amount of time a patient waits from the time he enters the department until his examination starts (WT1).
2. The amount of time a patient waits from the time a room is assigned until his examination starts (WT2).
3. Total time spent in the department (TT).
4. The number of patients in the department at the end of a working day (4:30 PM).

These measures might be different for inpatients and outpatients. One measure might be above another, but with different weights assigned for inpatients and outpatients. Also, different weights can be given to the different examinations (there are 13 major categories). For example, emhasis can be placed on trying to reduce the waiting times of those patients who require short examinations (10 minutes or less) rather than those who require lengthy studies (greater than 60 minutes).

Each scheduling procedure is evaluated on the basis of a combination of all the mentioned measures. Utilization figures cannot directly indicate performance, however, a reduction in waiting time, due to improved scheduling procedures, results in additional time available for processing patients. Since patient load is expected to increase, the ability to handle these additional patients increases the utilization of staff and equipment.

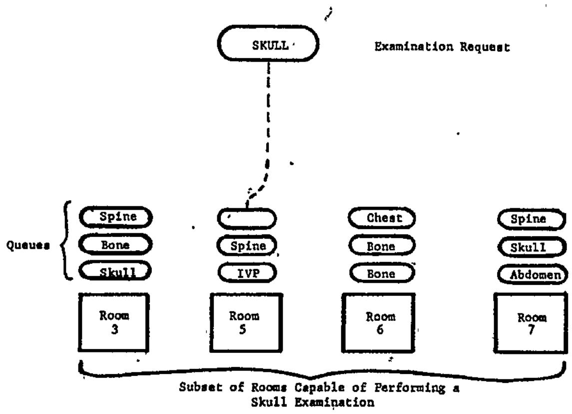
To evaluate the alternative scheduling procedures the simulation was run four times, each run representing a different patient mix. This was done to eliminate the effect of a particular mix on the solution. All other variables were held constant. Thus, variation in results is attributed to patient mix and scheduling procedures. If the procedures perform consistently over all mixes, than the average statistics of the four runs can be compared to determined which procedure produces the best results. If one procedure is not consistent, then to determine which is best requires the administrator to decide the order of importance for the various measures of performance. Weights can be assigned to each measure, and with these weights the best scheduling procedure can be selected.

IV. SCHEDULING ALTERNATIVES

SCHEDULE I - PS

The present scheduling system (PS) was used to compare the various alternative procedures. There are 13 major examination categories, each of which can be served by a particular subset of examination rooms. Once a subset is chosen, a decision is made at the control desk to determine which room a patient will enter. The decision is based on the number of patients waiting for service in front of each room contained in the subset. These queues are represented by request cards that are placed on an assignment board behind the control desk. As the day progresses, the board is updated and assignments are made accordingly. A schematic diagram depicting the queueing process is as follows:

ILLUSTRATION 2.



Such a process does not give a true picture of the amount of work in front of each room. For example, in the schematic room 5 would be selected despite the fact that an IVP examination is contained in the queue. This particular examination is a lengthy one taking approximately 60 minutes. Room 3 or 6 would be a better choice because the queues contain relatively short examinations. For this reason, patients experience excessive waits that could be eliminated if the work load in each queue was taken into account.

Schedule II MWL

The first alternative scheduling procedure tested considered the queue mix and work load when assigning rooms. Given a subset of rooms, a decision is made based upon the expected amount of work in front of each room. The queueing process is similar to the present system shown in Illustration 2. Instead of recording the number of patients waiting for a room, however, the amount of work in front of each room is noted. As a patient joins a particular queue his mean processing time is added to a load vector that accumulates the mean or expected work load for each room. This load vector is updated as patients enter and leave examination rooms. The patient is assigned to the room with the minimum work load (MWL).

The MWL policy takes into consideration the queue mix, and thus one can expect better results with regard to waiting time. What is not taken into account, however, is the amount of processing time remaining for the patient currently in the room at the time a room assignment is being made. The following example will clarify this point. Suppose a decision has to be made between two rooms whose work loads are 50 and 30 minutes respectively. Under the MWL policy the room with 30 minutes of work in the queue is selected, however, of this 30 minutes the patient currently being processed accounts for 20 minutes. If this patient entered the room 5 minutes ago, he has 15 minutes of processing remaining. Therefore, the true amount of waiting time for a patient joining this queue is 10 + 15 = 25 minutes. Now consider the 50 minute queue. Of this 50 minutes, the patient in the room accounts for 45 minutes. If he entered the room 35 minutes earlier he has 10 minutes of processing time remaining. The true amount of waiting time in this case is 5 + 10 = 15 minutes. The 50 minute queue should be selected to minimize waiting time. Under the MWL policy it is not selected because the remaining processing time of the patient in the room is not taken into account. As a result, even if the patient in the room only has a few minutes of processing remaining, his entire mean processing time is represented in the load vector until the patient leaves the examination room.

Schedule III - MMWL

The MWL rule is adjusted to include remaining processing time of the patient currently being served at the time a room assignment is being made. This rule is referred to as the modified minimum work load (MMWL). The decision variable in this policy can be written mathematically as follows:

$$\text{Min} \quad W_i = L_i - P_i + [\ P_i - (A - S)\ ]^+ \quad (1)$$

where:

- i is the set of all rooms that can perform this examination.
- L_i = the expected work load of each queue i for $i = 1,2,\ldots n$
- P_i = mean processing time of the patient currently in room i
- A = current time of the patient requiring a room assignment
- S = examination starting time of the patient currently in room i

If the term $[P_i - (A\text{-}S)]^+$, representing remaining processing time of the patient currently being served, is negative, then it is considered to be 0 for decision purposes, and just the mean processing time is subtracted from the load vector. This adjustment is necessary to allow for those times when an examination exceeds its expected duration (P_i). (In the simulation model the expected processing times are random variables which frequently exceed the expected value).

Schedule IV - CQ

In all of the above scheduling procedures patients joined multiple queues based on examination type and queue size. Given an examination category the patient is directed to a particular subset of rooms capable of performing the examination. The patient is assigned to the room that has the smallest queue. Once assigned, the patient remains in the same queue until his processing is complete. This procedure eliminates the possibility of patients being assigned to a different room if one happens to become available. Theorectically, service would be improved if all patients were served from one common queue (CQ). To test this procedure all patients who could not immediately be served upon leaving the control desk were placed on a waiting list. They are placed on the list according to departure time from the control desk. As rooms become available, the waiting list is scanned to find the first patient whose examination requirements can be met by the freed room. Thus, the order of selection is first-come-first-served (FCFS) as patients leave the control desk. This common queue policy allows for the interchangeability of rooms, and patients are not restricted to one room if another becomes available.

Schedule V - MCQ

A modification of the CQ rule was tried in order to minimize total time spent in the department. As previously mentioned, input to the department consists of inpatients and outpatients. Inpatients arrival time is recorded when they report to the control desk. Total time for inpatients is the interval between arrival at the control desk, and departure after processing. Outpatients arrival time is recorded when they report to the reception desk. After data is collected at the reception desk 80% of the outpatients have to dress for the examinations. Because of this preprocessing, outpatients usually report to the control desk 15 minutes after their arrival to the department. Consequently, outpatients join the waiting line for rooms behind inpatients who report directly to the control desk. As a result, the outpatient, who arrived earlier, is served after an inpatient who arrived at a later point in time. This procedure (CQ) tends to increase the total time of outpatients. To change this the CQ rule is modified by selecting from the queue according to arrival time. This rule gives priority to those patients who arrived earlier to the department, but were delayed prior to reporting to the control desk.

Schedule VI - SPT

As can be seen, the sequencing of patients through the system consists of two separate decision processes, 1) the selection of a room in which to perform the examination, and 2) the selection of patients from their respective queues. Room selection has been based on the number of patients in each queue and the amount of work in each queue. Patients for the most part have been selected from the queues on a first-come-first-served (FCFS) basis. The PS and MWL rules required patients to join multiple queues on a FCFS basis, while the CQ rules had patients join a single queue on a FCFS rule. To reduce average waiting time of all patients the shortest-processing-time (SPT) discipline is evaluated. Instead of selecting patients on a FCFS basis, priority is given to those examinations with small mean processing times. Highest priority is assigned to those examinations with the smallest mean processing time. SPT is reported in the literature as a good scheduling policy to minimize average waiting time and total time in a system (1).

Schedule VII - TSPT

When applying the SPT rule, the lengthy examinations run the risk of constantly being preempted by shorter ones. This may result in excessive delays for those patients whose examinations require large amounts of processing time. To avoid this, the SPT rule is modified by a truncation process. A critical limit for waiting time is established. Each time a room becomes available the waiting list is searched to see if any patient's waiting time had reached this critical limit. If so, this patient is given top priority and selected from the queue. If the critical limit has not been reached by any one patient, the SPT selection process continues. The critical limit serves as a control on waiting time, it does not guarantee that all patients will be served in this amount of time.

V. EVALUATION OF SCHEDULING PROCEDURES

In evaluating the various scheduling procedures each measure of performance is first analyzed separately to determine which procedure produces the best results. After each is examined separately, a combined analysis is conducted to determine a best procedure for all measures of performance.

The first performance measure analyzed is patient waiting time. It can be evaluated in terms of mean waiting time for all patients and the distribution of these waiting times. When evaluating the various scheduling procedures we cannot simply look at mean figures because this alone does not indicate how the department is operating. The objective of this work is to reduce the waiting time of all patients and at the same time eliminate those patients who experience excessive delays (greater than 60 minutes). The second part of this objective requires a truncation of the waiting time distribution. If the mean waiting time remains the same while the tail of the distribution is eliminated (reduce the number of patients whose wait exceeds 60 minutes) we can conclude that an improvement has taken place. On the other hand, if the tail is eliminated and mean waiting time increases sharply, a judgmental decision has to be made to determine whether or not an improvement has occurred. Ideally, the mean will decrease and the distribution will be truncated. Total time spent in the department can be analyzed in the same fashion.

	PS	MCQ	CQ	MWL	MMWL	SPT	TSPT
WT1							
OP	31 33 32 37	29 24 26 41	34 30 30 35	28 25 27 32	24 25 27 28	40 29 29 41	32 28 26 36
IP	18 20 16 18	15 18 20 24	17 23 14 22	9 15 10 14	9 15 10 15	16 22 14 22	15 24 15 21
BOTH	23 23 22 24	20 19 22 29	23 24 20 26	16 17 16 20	14 17 16 19	24 24 20 28	21 25 19 26
WT2							
OP	- 16 15 19	10 7 9 23	15 12 14 17	- 8 10 14	6 8 9 10	21 12 11 23	14 11 9 19
IP	- 19 15 17	14 16 19 23	16 22 13 21	- 13 9 13	7 14 9 13	14 21 13 21	13 22 13 20
BOTH	- 18 15 18	13 14 15 23	15 19 13 20	- 12 9 13	7 12 9 12	17 19 12 22	13 20 12 20
TT							
OP	85 68 73 90	83 61 67 95	91 64 72 86	85 61 69 85	79 61 68 81	97 64 70 93	89 63 67 87
IP	60 71 54 61	57 70 62 72	63 77 54 72	52 70 48 57	50 67 49 59	59 75 52 67	59 75 53 75
BOTH	69 70 61 70	66 68 64 79	73 74 61 76	64 68 56 66	61 65 56 66	73 72 59 76	70 72 59 79
No. at 4:30PM	19 21 20 18	15 17 17 21	15 18 14 24	12 18 13 13	14 17 13 14	16 19 15 20	16 18 12 24

Table 1A Mean Values of Waiting Time, Total Time, and the Number of Patients in the System at 4:30PM. (Four simulation runs were made for each scheduling alternative.)

	PS	MCQ	CQ	MWL	MMWL	SPT	TSPT
WT1							
OP	33	30	32	28	26	35	31
IP	18	19	19	12	12	19	19
BOTH	23	23	23	18	17	24	23
WT2							
OP	17	12	15	11	8	17	13
IP	17	18	14	12	11	17	17
BOTH	17	16	17	11	10	18	16
TT							
OP	79	77	78	75	72	81	77
IP	62	65	67	57	56	63	66
BOTH	68	69	71	64	62	70	70
No. at 4:30PM	20	18	18	14	15	18	18

Table 1B Average Values Over Four Simulation Runs for the Measures of Performance in Table 1A.

Table 1 presents data concerning mean waiting time and mean total time in the department divided among inpatients and outpatients. Each number in the table is an average for the 160 patients representing one working day. Waiting time is tabulated in two ways. The first (WT1) measures the interval from when a patient arrives to the department until he starts his examination. This includes all preprocessing of outpatients. The second (WT2) excludes the preprocessing and only measures the interval from when a patient leaves the control desk until he starts his examination. The division of waiting time is necessary to note the effects of various scheduling procedures on outpatient waiting time.

For these measures of performance the MMWL procedure produced the best results. In the majority of cases this procedure resulted in a mean waiting time that was less than or equal to the mean time produced by the other scheduling alternatives. The reduction in WT2 (compared to the present system) over the 4 patient mixes ranged from 28% to 56% for all patients. The average reduction for all 4 simulation runs was 41%. This amounts to reducing waiting time 18.5 hours per working day. Since patient load is expected to increase this freed time can be used to process the additional workload. The MWL policy produced results similar to the MMWL in terms of mean performance measurements. Additional simulation runs are necessary to determine whether or not the observed differences are statistically significant. The MMWL and MWL procedures are clearly better than the other scheduling alternatives. The differences between the remaining 5 procedures are slight, and it would be difficult to say which, if any, is better than another.

Table 1 also includes data pertaining to the number of patients in the system at 4:30 PM. With regard to this measure, the MWL and MMWL policies again seem to be superior to the other scheduling alternatives. In one of the 4 simulations, however, the TSPT rule produced a better result.

In order to determine which procedure performed the best in terms of the distribution of waiting time (WT1), the number of patients who experienced waits in excess of 60 minutes was recorded. Results are shown in Table 2.

	PS	MCQ	CQ	MWL	MMWL	SPT	TSPT
Run 1	12	17	28	-	3	22	23
Run 2	18	20	24	16	18	21	23
Run 3	9	18	19	8	8	15	11
Run 4	25	34	28	14	15	23	22
$\bar{X}$	16	22	25	12	11	20	20

Table 2 Number of Patients Waiting Greater Than 60 Minutes to Receive Service

The MWL policy performed consistently better than the other alternatives over all 4 simulation runs. The MMWL policy produced results close to those of the MWL policy. When comparing the MWL and MMWL procedures with the present policy (PS) the differences are not as great as those observed in Table 1. Here it seems that the PS performance can be grouped with the MWL and MMWL procedures, and not with the remaining alternatives which was the case when analyzing mean waiting time and mean total time.

Another way of interpreting the distribution of waiting time is to note the number of patients who obtain service in a specified amount of time. If, for example, under the PS alternative 95% of all patients received service in 75 minutes or less, while under the MWL policy 95% received service in 67 minutes or less, we can conclude that the MWL policy is an improvement in terms of truncating the distribution of waiting time. Such results are presented in Table 3.

	PS	MCQ	CQ	MWL	MMWL	SPT	TSPT
Run 1	60	75	90	60	52	105	90
Run 2	75	90	105	67	67	105	105
Run 3	60	97	90	60	55	82	67
Run 4	82	109	109	70	75	111	90
$\bar{X}$	69	93	99	64	62	101	88

Table 3 Distribution of WT1 (Numbers in the table represent the amount of time in minutes within which 95% of all patients received service.)

The MMWL and MWL alternatives have a greater number of patients receiving service in a shorter amount of time. The distribution of total time was analyzed in the same fashion. Results are shown in Table 4.

	PS	MCQ	CQ	MWL	MMWL	SPT	TSPT
Run 1	160	160	190	180	170	200	200
Run 2	160	180	180	167	160	190	180
Run 3	160	167	160	160	155	160	160
Run 4	170	190	177	180	173	200	190
$\bar{X}$	163	174	177	172	165	188	183

Table 4 Distribution of TT (Numbers in the table represent the amount of time in minutes within which 95% of all patients were in the system.)

The PS produced the best results in terms of this performance measurement, however, differences between the PS and the MMWL and MWL are small. Further experiments are needed to determine which policy is statistically better.

One final evaluation of the scheduling procedures consisted of dividing the 13 major examination categories into classes according to their processing times. It may be that administrators want to place emphasis on those patients whose examinations are short and therefore should not experience long delays. The examinations were divided into 3 classes and the mean waiting time (WT1) recorded. Results are shown in Table 5.

	PS	MCQ	CQ	MWL	MMWL	SPT	TSPT
0-15 mins.	22	24	25	17	16	27	22
16-45 mins.	27	25	25	20	20	27	28
> 45 mins.	25	25	28	27	29	26	29

Table V Mean Values of WT1 Over Four Simulation Runs Divided into Categories According to Expected Processing Time

In the less than 15 minute category the MMWL and MWL procedures produced the best results. Again it was difficult to distinguish which is the better of the two. The remaining alternatives produced similar results with the exception of the SPT rule which clearly was the worst in this category. SPT in theory should produce the best mean waiting time, however, this result tends to diminish as the flexibility of machine (room) selection increases (3). This seems to be the case in the radiology department where examinations can be performed in several different examination rooms.

Similar results are found in the 16-45 minute category, however, the SPT rule cannot be declared the worst. In the greater than 45 minute category the differences are not as great. The MCQ and PS procedures produce the best results. The data in this category are close enough to warrant further simulation runs and statistical analysis to determine the best procedure.

Based on the results presented in the previous tables, each scheduling procedure was ranked for the various measures of performance. These are listed in Table 6.

Rank	Waiting Time	No. at 4:30PM	> 60 mins.	95% Completed Service	> 45 mins.	0-15mins.	16-45mins.
1	MMWL	MWL	MWL	MMWL	MCQ	MMWL	MWL
2	MWL	MMWL	MMWL	MWL	PS	MWL	MMWL
3	MCQ	CQ	PS	PS	SPT	TSPT	MCQ
4	SPT	SPT	SPT	TSPT	MWL	MCQ	CQ
5	CQ	MCQ	MCQ	MCQ	TSPT	PS	PS
6	PS	TSPT	TSPT	CQ	CQ	CQ	TSPT
7	TSPT	PS	CQ	SPT	MMWL	SPT	SPT

Table 6 Scheduling Alternatives Ranked for Each Measure of Performance

In all but one measure, the MWL and MMWL policies produced the best results. The differences between these two procedures are small. These two procedures showed marked improvement over the other 5 policies analyzed. Among the remaining 5 policies no clear ranking over all measures was evident. Additional experiments with different patient mixes and loads may indicate which is better, but no statement can be made at this point.

When evaluating the various procedures, attention has to be given to the feasibility and cost of implementation. The CQ and MCQ procedures can be implemented manually but would require a considerable amount of bookkeeping and a means of scanning the entire queue to see which types of examinations are waiting. The patient whose characteristics match the available room would be selected. The SPT and TSPT procedures also could be implemented manually with the TSPT rule requiring some way to scan the queue for those patients whose waiting time exceeded the critical limit. The MWL and MMWL rules would require a computer to constantly update the expected load vector. Since the analysis showed little difference between these two, the MWL policy should be the one selected since it requires a simpler program to execute. A decision has to be made to determine whether or not the reduction in waiting time brought about by the MWL and MMWL procedures is worth the cost of implementation.

BIBLIOGRAPHY

1. Conway, R.W., et al., Theory of Scheduling Addison-Wesley, 1967.
2. Lev, B., Caltagirone, R., and Shea, F.J.: "Patient Flow and Utilization of Resources in a Diagnostic Radiology Department: Analysis by Simulation Techniques", Investigative Radiology, 7:517, 1972.
3. Wayson, R.D., "The Effects of Alternate Machines on Two Priority Dispatching Disciplines in the General Job Shop," Master's Thesis, Cornell University, February, 1965.