Patient Waiting Time System Using Simulation Method

Thet Sandar Aung
University of Computer Studies(monywa)
thetsandar2007@gmail.com

Abstract

This paper presents the way to calculate waiting times of patients. Results of patient surveys generally cite access to care and delays in care as sources of dissatisfaction. Delays are expensive, not only in terms of the direct costs they incur, but also in terms of the potential costs of decreased patient satisfaction and adverse clinical outcomes. In current health care service, it's important to minimize staff idle time and maintain a high utilization rate of the medical facilities. Computer simulation is a powerful tool for medical management and scheduling. In this simulation, patient waiting time is calculated step-by-step using queue theory. The main purpose of this paper is to reduce patient waiting time, period of time where the patient are waiting to cure or consult with doctor in clinics or medical care centers.

1. Introduction

Patient waiting time is a period of time where the patients are waiting to cure or consult with the doctor. An unplanned influx of patients entering a clinic will generally lead to poor service, patient congestion or queue if there are insufficient doctors or medical consultant running, or there is simply not enough medical equipments to serve the patients. Much of the time, it is impossible to predict and it is usually too late to so anything about it when it happens. Queues that form usually take a long time to disperse and this usually leads to patients frustration and dangerous.

This system is against step forward towards making the clinic or medical care center environment a more pleasant place to clinic. The system is designed as a tool to specifically solve these long time waiting issues. The system allows clinic owners to keep their finger on the "patient service" pulse, keep staffing at their optimum level and patients at their happiest and costs at or better than budgeted levels.

This paper examines the patient waiting and how to manage the queue. The discussion then reviews recent development in the market, services, customer service and relates these changes to understanding of waiting or queue systems.

Used of some general principles which, although not proven, have been reinforced over the years by personal experience and research investigations. Hopefully these principles will also be useful. They include:

- theory, simulation and practice should complement each other;
- systems integration is important;
- system commonality and structures are a useful basis for transferring experience;
- people are an integral and essential part of systems;
- automatic provision of data can greatly assist system maintenance;
- information is important;
- provision of queue management systems or information systems should be based on a proper economic assessment.

Many objectives must be considered in the design of a queue management technique. In particular, a clinic should consider fairness, safe, response time, service time, medical check up time, etc., some of these goals depends on the patient condition for example emergency case, normal cure, etc. So waiting time may affect the patient condition. Medical care centers or clinics need to arrange and co-ordinate every process stage related to queue the patients, consulting with medical consultants or doctors, taking medical checkup, curing safely.

There is a need to develop better ways of evaluating waiting time in appropriate capacity. The high rate of queue that has been described suggests that there is a need for non-priority or priority queues procedures that systematically adjust the system consisting of curing and patient waiting time so that they remain in safe.

2. Theory Background

A common situation occurring in everyday life is that of 'queuing' or waiting in a line. A queue is formed when either units requiring services – commonly referred to as customers, wait for service facilities, stand idle and wait for customers. An increase in the exiting service facilities would reduce the customer's waiting time and resultantly reduce queuing up. The basic of all servicing is the queue structure [5]. Queue servicing or scheduling is primarily used for serial processing.

There are two main types of queue.

(1)First-come first-serve (FCFS)

(2)Sorted queue, in which the elements (patients) are regularly ordered according to some rule. The most prevalent example of this is the priority queue.

The FCFS queue is the simplest and incurs almost no system overhead. The priority queue scheme can cost quite a lot in system overhead, since each process in the queue must be evaluated to determine which is highest priority one. The priority strategy is often used for important case to process since it is quite inexpensive to determine the emergency patients.

Processing time - The processing time represents the time taking has to spend on consulting.

Weight - The weight of process is a priority factor, reflecting the importance of process relative to other processes in the system. It may represent the emergency case in the system for one time unit.

Starting time -The starting time is the time when process starts its processing on consulting.

Completion time -The completion time is the time when process is completed on consulting.

Preemptive – Preemptions are less prevalent in services than in manufacturing. Many types of service activities are very difficult to preempt, e.g., an operation in a hospital, a flight leg or a game. However, some types of service jobs can be preempted, but usually such preemptions are not allowed to occur at just any time; preemptions in services tend to occur at specific points in time rather than at arbitrary points in time. The time during which a meeting takes place may be a collection of disjoint periods rather than one contiguous period. A meeting may be split up in segments of one or more hours each[6].

How queues behave depends on:

Arrival factors:

1. Arrival distribution -- When do new customers arrive? At random? In groups or singly?
2. Size of population from which customers are drawn -- Is the population effectively infinite or is it small enough that each arrival means one less new customer in the future?

Service factors:

1. Service time distribution -- how long it takes to serve an arrival
2. Number of parallel channels or servers, such as how many checkout counters at the grocery store
3. How many stages of service there are

Queue factors:

1. How many queues -- one or more than one?
2. Service priority among customers. Possibilities include

First come first served

Random

Customers with shorter expect service time go first

Triage and priority assignment, with or without preemption of customers already being served

3. Customer behavior in queue

balking or bounded -- customers don't join a line that's too long

reneging -- customers quit the line if the wait gets too long[2]

2.1 Service Facilities Model

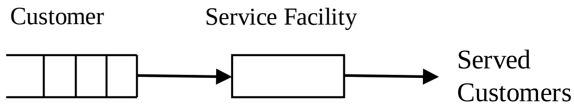

Figure1 Block diagram of single queue single service facility

Single-Server Single-Stage Queue

Arrivals → Queue → Service → Done

Several Single Server Single Stage Queues in Parallel

Arrivals → Queue → Service → Done

Arrivals → Queue → Service → Done

Arrivals → Queue → Service → Done

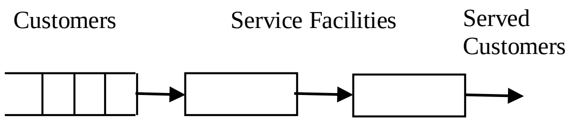

Figure2 Block diagram of single queue multiple service facilities

Single Server Multiple Stage Queue

Arrivals → Queue → Service → Service → Service ... → Done

The Multiple Stage model is like a cafeteria, in which customers waiting to get their main dishes may prevent customers behind them from getting their salads[2][7].

Single Server Queues in Series

Arrivals → Queue → Service → Queue → Service ... → Done

The Queues-in-Series model has queues between the stages of service, which the Multiple Stage model does not. This might be appropriate for a doctor's office. Patient waits in the waiting room until seen by a nurse. Then they wait again in examining rooms until the doctor arrives[2][7].

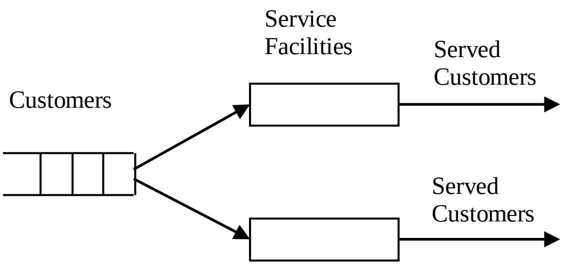

Figure3 Block diagram of single queue multiple service facilities in parallel

Multiple Server Single Stage Queue

```
                    ┌→ Service → Done
Arrivals → Queue ─┼→ Service → Done
                    └→ Service → Done
```

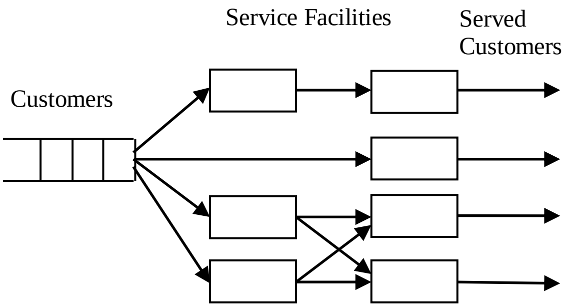

Figure4 Block diagram of single queue multiple service facilities in mix model

There have multiple service facilities in mix mode to serve the customers. When one customer is being served in one service facility, another customer is also being served in another service facility where customers are served any free service facilities with parallel mixing mode[2][7].

2.2 Queuing Theory

Queuing theory is the mathematics of waiting lines. It is extremely useful in predicting and evaluating system performance. Queuing theory has been used for operations research. Traditional queuing theory problems refer to customers visiting a store, analogous to requests arriving at a device. Solution of queue models are two approaches, namely mathematical and simulation, which are available for solving queuing models [1].

Long Term Averages: queuing theory provides long term average values. It does not predict when the next event will occur. Input data should be measured over an extended period of time. There are assumed arrival times and service times are random [4].

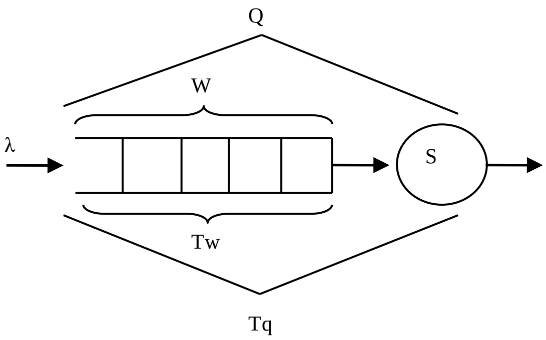

Figure5 Queuing model

The symbols and interesting values in figure5 as below:

- Arrival rate (λ) — the average **rate** at which customers arrive.
- Service time (s) — the average **time** required to service one customer.
- Number waiting (W) — the average **number** of customers waiting.
- Number in the system (Q) — the average total **number** of customers in the system.

The symbols and more interesting values are as follow:

- Time in the system (Tq) the average **time** each customer is in the system, both waiting and being serviced.
- Time waiting (Tw) the average **time** each customer waits in the queue.

$$Tq = Tw + s$$

Arrival Rate: The arrival rate, λ, is the average rate new customers arrive measured in arrivals per time period. Common units are access/second. The inter-arrival time, **a**, is the average time between customer arrivals. It is measured in time per customer. A common unit would be seconds/access [4].

$$a = 1 / \lambda$$

Random Values: Most of the events we are interested in occur randomly.

- Time of a request to a device
- Time to service a request
- Time user makes a request

Although events are random, we may know the average value of the times and their distribution.

Exponential Distribution: many of the random values are exponentially distributed.

$$\text{Frequency of Occurrence} = e^{-t}$$

There are many small values and a few large values. The inter-arrival time of customers is naturally exponentially distributed [4].

2.3 Common Models of Queue

The simplest queuing model is **M/M/1** where both the arrival time and service time are exponentially distributed.

The **M/D/1** model has exponentially distributed arrival times but fixed service time.

The **M/M/n** model has multiple servers.

A full queuing situation involves arrivals and service, so some more Greek letters are needed:

λ "lambda" is the average customer arrival rate per unit time.

μ "mu" is the average customer service rate (when customers are waiting).

μ = 1/(average service time)

ρ "rho" is λ / μ, also called the server utilization factor.

The Poisson and exponential distributions will used to model both arrival times and service times. As mentioned, the Poisson and exponential distributions are mathematically related. If the number of service completions per unit of time, when there is a backlog of customers waiting, has the Poisson distribution, then service time has the exponential distribution. It's conventional to think of service in terms of the length of service time.

The standard simple queuing model assumes that:

1. arrivals have the Poisson distribution
2. service times have the Exponential distribution
3. arrivals and service times are all independent

(Independence means, for example, that: arrivals don't come in groups, and the server does not work faster when the line is longer.)[4]

2.4 Single Server Models

Many systems give rise to single server models. Single server models are also important in decomposition methods, when scheduling problems in more complicated server environments are broken down into a number of smaller server machine scheduling problems.

Single server models have been thoroughly analyzed under all kinds of conditions and with many different objective functions. The result is a collection of rules that, while easy to identify and apply, often provide optimal solutions in the single server environment [6].

2.5 Parallel Server Models

A bank of servers in parallel is a generalization of the single server model. Many production or service environments consist of several stages, each with a number of servers in parallel. The servers at a "work center" may be identical, so that a job can be processed on any one of the servers available. Parallel server models are important for the same reason that single server models are important. At times, the servers in parallel may not be *exactly* identical. Some servers may be older and operate at a lower speed; or, some servers may be better maintained and capable of doing higher quality work. If that is the case, then some jobs may be processed on any one of the *server*, while other jobs may be processed only on specific subsets of the *server*. When the "servers" are people, then the processing time of an operation may depend on the job as well as on the person [6].

2.6 Parallel Processing

Parallel processing is the method of breaking large problems down into smaller constituent components, tasks or calculations that are solvable in parallel [9]. Parallel processing has emerged as a key enabling technology in modern computing. This has been a result of a demand for higher performance, lower cost, and sustained productivity. The acceptance of parallel processing has been facilitated by two major developments: massively parallel processors (MPPs) and the widespread use of distributed computing [6].

Advantages of parallel processing are that common architectural element between distributed computing and MPP's is the notion of message passing. In all parallel processing, data must be exchanged between cooperating tasks. In this area of research, several paradigms have been experimentally evaluated including shared memory, paralleling compilers, and message passing. The message-passing model has become the paradigm of choice, from the perspective of the number and variety of multiprocessors that support it, as well as in terms of applications, languages, and software systems that use it[6].

3. Proposed System

There have five models

- Single queue single service model
- Single queue multi-service sequence model
- Single queue multi-service parallel model
- Single queue multi-services multi-parallel model
- Single queue multi-services multi-parallel with auto arrangement model

"Single Queue Single Service Model" system will perform one after another. The service counter always busy until the queue is empty and the patients will take more time for waiting because single service model is performed individual patient till they have finished and completed.

"Single queue multi-service sequence model" system will perform with two service counters. First patient is servicing in service counter 1 and second patient is servicing in service counter 2. Because of the first patient has passed from service counter 1 and the patient go to service counter 2 continually the second patient can be served in service counter 1. When single service model is used, the patients are needed to wait till the one patient is completely finished but now the patients are not needed to wait till the one patient is completely finished, they just only need half of one patient taking time. So this service model can guess quicker than first service model. This model is safer for emergency patients than single service model.

"Single queue multi-service parallel model" will perform parallel with two service counters. Each service counter performs the patient until they have finished. This service model is quicker than single queue multi-service sequence model because of it can complete one time two patients so it can reduce the waiting time rather than single queue multi-service sequence model.

"Single queue multi-services multi parallel model" will perform multi parallel. This model must have minimum of 4 service counters. But it can operate two patients at a time where one service counter performs half of the patient actual service time on front counters and the remaining time will serve on back service counters. So it can actually serve four patients on current. But the drawback of this model is that the backhand service counters wait the patients from front hand counters. So it can have idle time on service counter. It is good for patient but it wages the time.

"Single queue multi-services multi-parallel with auto arrangement model or idle arrangement model" will also perform multi parallel and will go to auto. This model must also have minimum of 4 service counters. But it performs two patients at a time where one service counter performs half of the patient actual service time on front counters and the remaining time will serve on back service counters. But this model does not wage the time because of every service counter always monitors queue and receives the patient when they are free. So this model is the best model not only on the simulated system but also real time customer service facility counter.

The patients waiting time by using simulation method work as follow:

- the system received the patients into queue
- after adding the patients into queue, the system check the queue configuration that is whether the queue type priority queue or non-priority queue
- if the queue type priority queue sort the patients according to the priority level
- and then the system choose the service mode configuration
- if service mode is single service facility mode, the system serves the patient until the patient is completed finished
- if the service mode is multiple services sequence mode, the system serves a patient first service facility and then the patient is served in another service facility for second time where another service facility is already served another patient in queue, this means patients are served one after another
- if the service mode is multiple service parallel, the system serves the patient in one service facility until to finish and another service facility is already serve in another free service facility
- if the service mode is multiple service mix mode, the system serves the patient in any free service facility and wait to serve the patients when they are free state
- when then service facilities are free state, they always check the queue to serve any patient

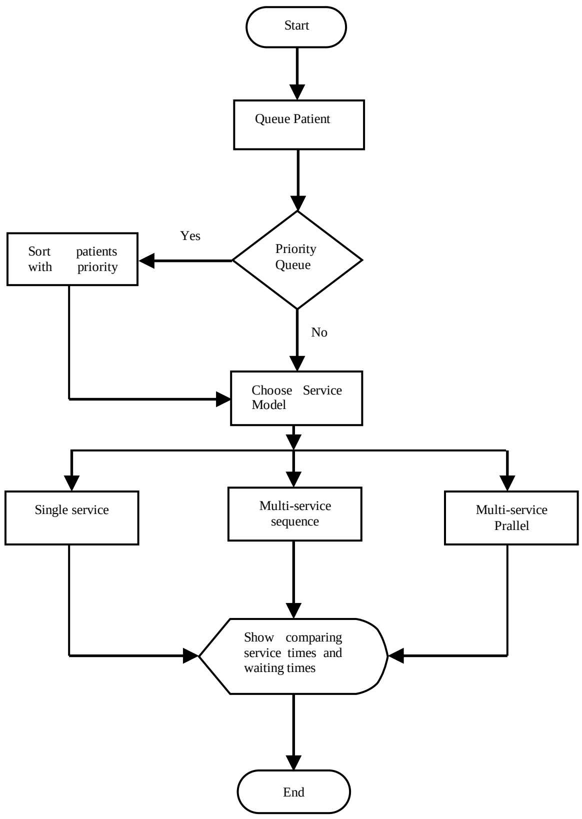

Figure6 System design

8. Simulation

The process of designing a model of a real system and conducting experiments with this model for the purpose either of understanding the behavior of the system or of evaluating various strategies (within the limits imposed by a criterion or set of criteria) for the operation of the system[1].

Simulation is a useful or powerful tool for analyzing complex systems and used in order to study real-life systems which do not currently exist [3]. Good for representing complex systems, the utility of the technology for finding the “answer”. Easy to learn and are applicable to a wide range of problems.

When everything else fails, then simulate. Any real-life system studied by simulation techniques is viewed as a system. When building a simulation model of a real-life system under investigation, one does not simulate the whole system [5].

The simulation theory is an account of everyday ability to make sense of the behavior of others. The successful ‘mindreading’ of mental states allows us to predict and to explain what others do, and makes possible the rich social dynamic that pervades human life.

9. Simulation Result

This form is the service compared chart that the result of patient waiting time after processing. Five service models can compare on this chart don’t considering queue policy where priority or non-priority because of priority does not affect for service just only for patient.

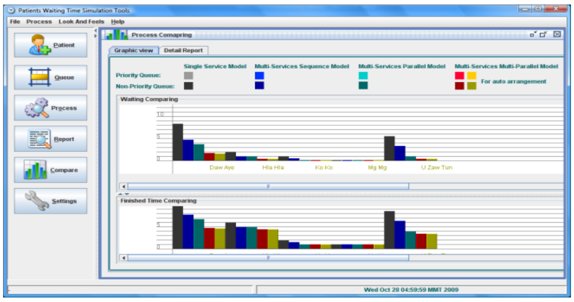

Figure7 Simulation compared bar chart

In this page can see the details of comparative text report. This report can shows the average waiting and average done time for each service model. Text report is more precise than bar chart report. So, for each service model can compared carefully.

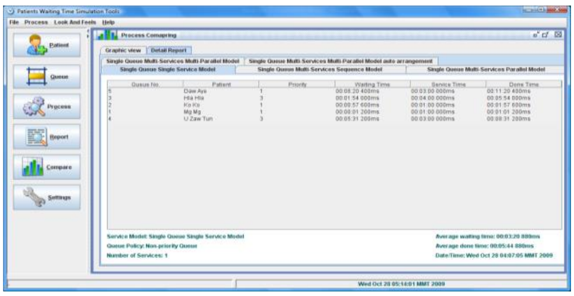

Figure8 Simulation comparisons with text report

10. Conclusion

This system presents the patient waiting time by using simulation technique and queue management theory. As simulation technology is applied to increasingly complex systems, traditional methods of output reporting may not provide sufficient information to decision makers. The simulation results show that the maximum queue contents. Waiting time of the patients increases drastically. The purpose is to reduce patient waiting time. This system can compare the service facilities with their average waiting time and average done time. Besides, the system simulates about the queue type and policy, it shows how the priority level is important for patient waiting system. The main purpose is to investigate how to control waiting problems.

Concept, design and implementation have been also shown of this case study, Patients Waiting Time System**,** in which a real-world waiting time is simulated on a simulation system using Priority and Non-Priority Queue.

11. References

[1] B.Ashish, “Queuing Theory Model For Management Of Semiconductor Fabs And Test Floors For Optimal Utilization Of The Equipment”

[2] Samuel L. Baker, “Queuing Theory I”, 2006

[3] Dr. Hans Kraml “Simulation Theory versus Theory Theory”, Universit¨at Innsbruck eingereicht bei Herrn Univ Innsbruck, M¨arz 2002

[4] KenW, “Queue Theory”.

[5] Harry Perros, “Computer Simulation Techniques:The definitive introduction”, Computer Science Department NC State University Raleigh, NC

[6] Michael L. Pinedo, “Planning and Scheduling in Manufacturing and Services”, Department of Operations Management, Ster School of Business, New York University, 40 West 4th Street, Suite 700, New York, NY 10012-1118, USA, mpinedo@stern.nyu.edu, ISBN 0-387-22198-0, ©2005 Springer Science+Business Media, Inc.

[7] Jk Sharma, “Operations Research Theory And Applications”