Istanbul Technical University

Graduate School of Science Engineering and Technology

Computer Engineering

BLG 517E Modelling and Performance Analysis of Networks

2016 – 2017, Spring Semester

From computer science to service science: Queues with human customers and servers by Hideaki Takagi

504161508 Damla Gülen

Professor Kleinrock has made significant contributions to the computer communication networks. In this paper one of his works, queueing theory is taken into consideration. Queueing theory is studied in the field of service systems having human customers and servers. The main focus is on the potential of the queueing theory in the scope of emerging *service science*.

1. Background information about the author

The author of the paper is one of Professor Kleinrock's students. Hideaki Takagi, worked on the application of queueing theory to computers and communications networks and recently interested in mathematical and statistical treatment of service systems with human customers and servers in the context of an emerging discipline called service science. According to Takagi, service science with human customers and servers is one of the fields where queueing theory can still be effective.

2. Service systems with human servers and facilities

Service can be defined as the activity to bring value (satisfaction) to not only the recipients (customers) but also the providers (employees) by optimal management of a set of available resources. There are services without human servers such as online reservation services or Internet banking applications. However, many services still involve human servers such as counters at airport or service at a restaurant. In services that involve human servers, employers provide service by using some physical facilities as enablers. The service is different from goods in terms of intangibility which is the difference of being able to be physically touched. In these systems, the operational resources are employees and facilities which are not free and require investment. In order to administrate the service, manager holds a finite number of employees and facilities to supply them when demanded.

A basic flow of a service organization is as follows: Customers come to the service where there are certain number of resources allocated to services for a certain duration of time. Customers are served by the appropriate servers. If there is no available server at the time when the customers arrive, they are forced to wait in a *queue*. When there is no available space in the queue, the customers are not admitted to the system anymore, until there is space in the queue or a server is free. When a server becomes available, mostly head of the queue is selected for the next service. The importance of staff scheduling and facility scheduling are necessary to both satisfy the customer and employee needs. In this paper, a real time example of a hospital is studied. In hospitals, patients are customers and medical equipment, surgery rooms, beds are facilities and doctors, nurses, and technicians are the servers. Different kinds of services can be surgery, examination, diagnosis, clinical treatment and rehabilitation.

When there is insufficient resources the organization works with delay, which drags the system into the *congestion state*, where waiting time gradually increases since there is less servers than the demand. Customers who wait longer in the queue are tend to complain more about the service when they interact with the (human) server. This results in longer service time and waiting time. This situation mostly effects the servers, and they tend to get upset which results in bad quality of service and slower service time. Job satisfaction of the employees and customer satisfaction are needed to be consider in these systems since customer and employee interact directly they both effect each other's attitude which effects systems service rate. Satisfaction of servers is different from the other queueing modelled systems.

3. “Power” of Kleinrock

In service systems managers aim to keep his resources fully utilized by accepting as many customers as possible, but this case is not optimal for the customers. They demand shorter waiting queues and faster service. This situation is called trade-off of the resource utilization and customer satisfaction. Controlling customer arriving process is the key in order to have neither a crowded nor too empty system. A system performance metric called *power* aims to find the optimal level of controlling the customer arrivals.

In a stable state, where input rate equals the output rate, and T being mean response time the power ratio is defined as follows:

$$P(\lambda) := \frac{\lambda}{T(\lambda)}.$$

For M/M/1 queues:

$$T(\lambda) = \frac{1}{\mu - \lambda},$$

therefore,

$$P(\lambda) = \lambda(\mu - \lambda),$$

When $\lambda = \mu/2$, $P(\lambda)$ has the highest value which is unity (1).
For M/M/m queues, $P(\lambda)$ has the highest value when λ/μ ratio is $2/\sqrt{3}$. Then we have $E[N] = \sqrt{3} = 1.732$… These values show that there should be exactly one customer per server on average at the optimal operating point.

4. Patient flow in a hospital

In this section an example of queues with human servers and facilities is studied with patient flow in a hospital. Basic performance measures are as follows: bed utilization and length-of-stay (LOS) of each inpatient. Bed utilization is the ratio of the mean number of patients staying in the entire hospital or in each clinical ward to the total number of beds that the facility has. A complete data set of every movement of all the inpatients from room to room covering two years was provided by Medical Information Department of the University of Tsukuba Hospital (UTH) in Japan. The data set from obstetric unit of UTH is studied.

4.1. Little’s law applied to inpatients

In the obstetric unit of UTH there are two different treatment types; one for normal and one for high-risk childbirth. There are two wards numbered 30 M and 300. 30 M has 6 beds and is the maternal and fetal intensive care unit (MFICU), for the treatment of high-risk delivery. Ward 300 has 26 beds which accommodate patients with normal delivery and is a waiting room (queue) for Ward 30 M. Calculated and given values can be seen on table 3 in the original paper. There were 731 nights in the time period that the data set belongs to. The mean number of patients staying in bed on each night is calculated as 12.630/731 which is 17.28 patients/day, which leads to the bed utilization of 0.665 (=17.28/26). *LOS* for each patient is 6.43 days.
LOS of an arbitrary patient (W) can be calculated from the mean number of patients being hospitalized each day during a given observation period (L) and the number of new patients admitted per day in the same period (λ). This is done by applying Little’s Law: $L = \lambda W$. The only condition for this law is that the system is stable in that the number of patients present in the system does not grow indefinitely. When Little’s law is applied $W = 17.28 / 2.685 = 6.436$ which virtually agrees with the reported value $W = 6.43$ days as it should.

4.2. Queueing network model of the obstetric patient flow

In this subsection 5 scenarios of child-births are studied which can be seen on the figure 6 on the original paper. A queueing model of M/G/∞ and M/M/m queues because of its simplicity in modelling and computation.

4.3. Distribution for the number of patients in each ward

It was observed that the probability of all the beds are occupied is only about 1% for Ward 300, therefore, it is assumed that there are sufficient number of beds in Ward 300 to accept all patients at any time (M/G/∞) and Ward 30 M uses M/M/6 queue. M/G/∞ system has sufficiently many servers to which customers arrive in a Poisson process and spend there a random amount of service time which is generally distributed probabilistically. If λ denotes the arrival rate and *b* denotes the mean service time, the number N of customers present in the M/G/∞ queue at an arbitrary time has the Poisson distribution with mean *p*= λ*b.*

$$P\{N = k\} = \frac{\rho^k}{k!}e^{-\rho} \qquad k \geqslant 0.$$

With the given arrival and service rates, possibilities for different routes are calculated, for more detail the original paper can be seen. For probabilities of ward 30 M, please refer to the equations 11, 12, and 13 in Takagi's article with the results presented in figure 7 and 8. Theoretical results match with the major characteristics of observed values, but there is a remaining challenge for better agreement. The modelling and analysis of the obstetric patient flow is concluded with this remark.

5. Research of queues in human service systems

From this study it can be clearly seen that *perception of customers* should be considered. Other challenges that queueing theorists need to consider in human service systems are as follows;

Many servers because human service system mostly provides service with many servers on mediocre speed which brings less waiting time than one high-speed server. Time-varying customer arrival rate, because in human service systems the arrival rate of the customers change usually in different times of the day or year. The theory is insufficient in this field. Satisfaction of customers and servers need to be covered. Until now the focus was on the customer satisfaction but since this is a human service system where both customer and server effect each other's happiness they both need to be satisfied by meeting their demands. Workforce management is also an important challenge since the customers and servers are interacting during the service, unsatisfied employees tend to have negative impact on the customers with longer service time and bad quality of service.

6. Conclusion

The properties, evaluation metrics and challenges of a service system with human resources is studied with the example of a patient flow in a hospital. Patient flow was modelled and analyzed. For any detailed calculations and results Takagi's article listed in references can be referred to.

7. References

- H. Takagi, From Computer Science to Service Science: Queues with Human Customers and Servers, *Elsevier Computer Networks,* vol. 66, pp. 102-111, June 2014.