ISSN 2091-1521 (Print)
2091-153X (Online)
International Journal of Operational Research/Nepal - IJORN - 2016, Vol. 5, Issue 1

Brief Literature Review of the Queuing Problem

Yogesh Shukla[1] & Dr R.K. Shrivastav[2]
[1]AMITY University Madhya Pradesh, Gwalior
[2]SMS Science College, Gwalior
e-mail:yogeshparashar4jan@gmail.com, r.k.shri19@gmail.com

Abstract

Queuing problem is an important problem in our daily life. In general it is about waiting line to take any type of service in any type of store. Operations research plays a big role in solving such problems. In this paper a general systematic approach towards studying the literature of the Queuing model is considered which has been recently used by a number of researchers in the field of mathematics and management. In addition to that, we aim to identify promising future works from the study results. The review process is based on our variation of an existing literature review method. This variation is also presented in the paper. Although a concern in any type of store, Queuing model has not been widely studied when compared to other similar fields of research. It has been mainly studied in operations research and in the context of machine repair problem. It seems that artificial intelligence and machine learning still have a good potential to contribute to this field of research in different applications. Application of Queuing problem in Mathematics and Management is an open area of research. For instance, it looks promising for developer in the field of mathematics and management.

Keywords: Queuing problem, Queuing model, Systematic literature review

1. Introduction

Queuing model is a broad area under the operation research in mathematics and management also. Queues (or waiting lines) help facilities or businesses provide service in an orderly fashion. Forming a queue being a social phenomenon, it is beneficial to the society if it can be managed so that both the unit that waits and the one that serves get the most benefit. For instance, there was a time when in airline terminals passengers formed separate queues in front of check-in counters. But now we see invariably only one line feeding into several counters. This is because of the realization that a single line policy serves better for the passengers as well as the airline management. Such a conclusion has come from analyzing the mode by which a queue is formed and the service is provided. The analysis is based on building a mathematical model representing the process of arrival of passengers who join the queue, the rules by which they are allowed into service, and the time it takes to serve the passengers.

2. Literature Review

Queuing theory's history goes back nearly 100 years. Johannsen's "Waiting Times and Number of Calls" seems to be the first paper on this subject. But the method used in this paper was not mathematically exact and therefore, from the point of view of exact treatment, the paper that has historic importance is A. K. Erlang's[1], "The Theory of Probabilities and Telephone Conversations" .Queuing theory embodies the full gamut of such models covering all perceivable systems which incorporate characteristics of a queue. Queuing theory, as such, was developed to provide mathematical models to predict behavior of systems that attempt to provide service for randomly arising demands and can trace its origins back to a pioneer investigator. Work continued in the area of telephone applications, and although the early work in queuing theory picked up momentum rather slowly, the trend began to change in the 1950s when the pace quickened and the application areas broadened well beyond telephone systems. Approximations have appeared in the literature. For an extensive bibliography Bhat et al [4] to mention a few, one approach to approximation is the analysis under heavy traffic (when the traffic intensity, the ratio of the rates of input to output, approaches 1) and investigations under this topic were initiated by Kingman (for an extensive bibliography, Kingman [10]) with the objective of deriving a simpler expression for the final result. The heavy traffic assumption also led to diffusion approximation as well as weak convergence results by researchers such as Iglehart [9]. Gaver's analysis [8] of the virtual waiting time of an $M/G/1$ queue is one of the initial efforts using diffusion approximation for a queuing system. Fluid approximation, as suggested by Newell [16,17] considers the arrival and departure processes in the system as a fluid flowing in and out of a reservoir, and their properties are derived using applied mathematical techniques. For a recent survey of some fluid models see Kulkarni [12].

3. Research Methodology

This section presents the research methodology (RM).which we have adopted. We start with the research questions which initiated and guided the process of literature review. Then we explain the way the list of search terms was identified and refined.

The investigated research questions (RQs) are:

RQ1: What are the solution and the validation methods in Queuing system literature?

RQ2: How much activity has been there on Queuing system?

RQ3: What are some of the limitations of the current research?

RQ4: What are some of the white spaces (potential future works) in this field of research?

The phases of our research methodology are shown in Fig. 1.

The literature review started with identifying the scope of the research and then preparing a primitive representing set of keywords and terms. This step is represented by the first (and the largest) outer layer in Fig. 1, where the size of a layer implies the relative size of the search space in its corresponding phase.

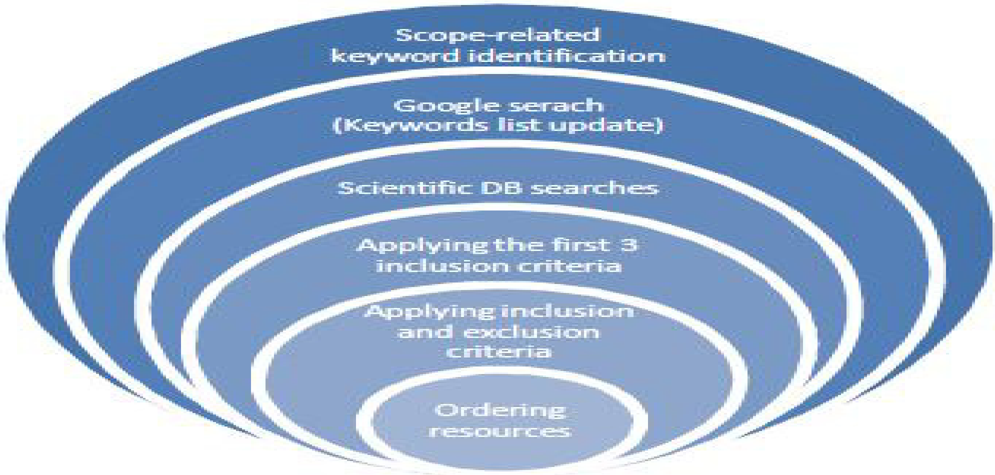

Fig.1: Phases in our literature review research methodology.

These initial set of keywords, terms and their combinations were used to search for the related work in Google. Although Google search results might not be purely scientific, they are beneficial in the sense that they provide a good insight into the commonly used field-specific terms and keywords. Thus, they help to refine the initial pool of terms and enable better searches to follow in scientific databases. This could be especially advantageous to researchers who are new to a field of research. The refined set of terms along with their combinations was looked up in the following major online scientific databases (DBs) and search engines: Google Scholar, IEEE Xplore, Scopus, ProQuest, ACM Digital Library, and Science Direct.

And finally we conclude that the simplest form of queuing models are based on the birth and death process, where the birth process describes the inter-arrival time (time between two arrivals) to the queue and the death process describes the service or holding time in the queue. Queuing theory exhibits the memorylessness property which often denotes a Markovian property and a process with a Markovian property is called a Markov process, which means that the probability distribution of future states of the process, given the present state and all past states, depends only upon the present state and not on any past states. As a result, queuing models are frequently modeled as Poisson processes through the use of the exponential distribution. Suppose that the inter-arrival time is described by an exponential distribution with parameter λ(traffic intensity), and the holding time is described by an exponential distribution with parameterμ. Then the transient behavior of the queuing system is expressed by:

$$P'_i(t) = \lambda P_{i-1}(t) + (\lambda + \mu)P_i(t) + \mu P_i(t) \qquad (1)$$

where $P'_i(t)$ is a derivative to $P_i(t)$which is the probability to have *i* in the queue system at time *t*. The system is described as a function of time and can be solved when we know the starting value at time 0.Suppose that the system reaches statistical equilibrium. Then the solution is independent of the starting values. In addition, it has a balance between interarrivals and services which implies λ/μ< 1. Then $P'_i(t)$= 0. Letting$P_i(t)$) =P_i, we get:

$$(\lambda + \mu)\, P_i = \lambda\, P_{i-1}, \; + \mu\, P_{i+1}, \tag{2}$$

Which is identical to the Gambler's Ruin Problem?

This queuing system is denoted M/M/1: Exponential inter-arrival time and holding time and one server.

The classification of queuing systems follows Kendell's definition . The solution is found by expressing all the $\{P_i\}$ as a function of P_0 and then normalize based on the summing up of all the probabilities to 1. The same procedure is done for the Gambler's Ruin problem, but the edge conditions are also taken into account. To show the equality in the solutions the following notations are used:

$$p(i) = P_i \tag{3}$$

$$\rho = \frac{\lambda}{\mu} = \text{p/q} \tag{4}$$

$$K = n + m \tag{5}$$

Solutions of different queuing systems and the Gambler's Ruin problem:

$$p\,(i) = (1 - \rho^i) / (1 - \rho^K) \tag{6}$$

$$\text{M/M/1:} \quad p\,(i) = (1 - \rho)\, \rho i \tag{7}$$

$$\text{M/M/1/K:} \quad p(i) = \rho^i\, (1 - \rho) / (1 - \rho^{K+1}) \tag{8}$$

$$\text{M/M/}\infty: \; p(i) = (\rho^i / i!)\, e^{-\rho} \tag{9}$$

$$\text{M/M/K/K:} \; p(i) = {(\text{pi/i!})} \Big/ {\textstyle\sum_{k=1}^{K}} \rho^k/k! \tag{10}$$

Erlang's B loss formula:

$$p(k) = {(\rho^k/k!)} \Big/ {\sum_{k-1}^{K} \rho^k/k!} \tag{11}$$

Here, A/B/C/D follows the notation of Kendell and Kleinrock, [11] where:

A: Interarrivals time distribution

B: Service time distribution

C: Number of servers

D: Waiting room capacity

It could be noted that the solution of queuing models with more than one server uses $(i+1)\mu$ instead of μ in equation (2).It is interesting to note that most books in queuing theory use equation (2) as a standard formula because it is derived from the transient equation (1). The stationary state equations can be interpreted as follows: The traffic stream out of state i is equal to the traffic stream into state i. Looking at the original work of Erlang , Brockmeyer [6] uses another approach. Instead of assuming that the traffic stream out of a state is equal to the traffic stream into the state, he postulates that the traffic stream is equal both ways between a cut of states. He then gets the simplified equation:

$$\lambda p(i)=(i+1)\mu p(i+1) \tag{12}$$

Where he also uses $(i + 1)\,\mu$ as a more general expression.Arne Jensen uses the same approach in his paper . A complete proof is given in Morris [13]. The possibility of using the cut is a much more powerful approach for modeling more complicated stationary queuing systems.

4. Conclusion

In the preceding paragraphs we have outlined the growth of queuing theory identifying major developments and directions. For details of any of the facts readers are referred to the articles and books cited above. Also see Prabhu [18] who gives a bibliography of books and survey papers in various categories and subtopics, Adan et al. [2] who give a broad treatment of queues with multiple waiting lines, and Dshalalow[7] who considers systems with state dependent parameters. The last two articles also provide extensive bibliographies. It is hoped that with the help of these references applied researchers will be able to build on the systems covered in this text so as to establish an appropriate model to represent the system of their interest.

References

[1] A. K. Erlang's (1909) "The Theory of Probabilities and Telephone Conversations" (*Nyt tidsskrift for Mathematic, B, 20* , p. 33

[2] Adan, I.J.B.F. Adam, Boxma, O. J. and Resing, J. A. C. (2001). "Queuing models with multiple waiting lines", Queuing Systems **37**, 65-98.

[3] Armero, C. (1994). "Bayesian inference in Markovian queues", Queuing Systems **15**, 419-426.

[4] Bhat, U. N., Shalaby, M. and Fischer, M. J. (1979). "Approximation techniques in the solution of Queuing problems", Nav. Res. Logist. Quart.,**26**, 311-326.

[5] Bhat, U. N. (1969). "Sixty years of Queuing theory", Management Sci.,**15** B, 280-294.

[6] Brockmeyer, E., Halstrom, H. L. and Jensen, A. (1960). The Life and Works of A. K. Erlang, Acta Polytechnica Scandinavica, Applied Math. and Comp. Machinery Series No. 6, Copenhagen.

[7] Dshalalow, J. H. (1997). "Queuing systems with state dependent parameters", Frontiers in Queuing (Ed. J. H. Dshalalow), Ch. 4, CRC Press, New York, 61-116.

[8] Gaver, D. P. Jr. (1968). "Diffusion approximations and models for certain congestion problems", J. Appl. Prob. **5**, 607-623.

[9] Iglehart, D. L. and Whitt, W. (1970). "Multiple channel queues in heavy traffic I", Adv. Appl. Prob., **2**, 150-177.

[10] Kingman, J. F. C. (1965)."Heavy traffic approximations in the theory of queues",Proc. Symp Congestion Theory (Eds.W. L. Smith and W.E.Wilkinson),Univ. of North Carolina,Chapel Hill,137- 169.

[11] Klinrock. L. (1975). Queuing Systems Vol. I: Theory, John Wiley & Sons,New York.

[12] Kulkarni, V. G. and Liang, H. M. (1997). "Retrial queues revisited", Frontiers in Queuing (Ed. J. H. Dshalalow), CRC Press, New York, Ch. 2,19-34.

[13] Morse, P.M. (1958). Queues, Inventories and Maintenance, John Wiley & Sons, New York.

[14] Neuts, M. F. (1975)."Probability distributions of phase type", Liber Amicorum Prof. Emeritus H. Florin, Department of mathematics, Belgium: University of Louvain, Belgium, 173-206.

[15] Neuts, M. F.(1978)."Markov chains with applications in Queuing theory, which have a matrix geometric invariant probability vector", Adv. Appl. Prob. **10**, 185-212.

[16] Newel, G. F. (1968). "Queues with time dependent arrival rates I - III", J.Appl. Prob. **5**, 436-451, 579-606.

[17] Newel, G. F. (1971). Applications of Queuing Theory (Ch. 6), Chapmanand Hall, London.

[18] Prabhu, N. U. (1987). "A bibliography of books and survey papers on Queuing systems:Theory and applications", Queuing Systems, **2**, 393-398.