Proceeding of 3"* /nfernaf!OMa/ Science Po^gro^Mafe Cow/erence 2075(ZS'PC2073)
6 Faculty of Science, UniversitiTeknologi Malaysia

TRANSIENT ANALYSIS OF M/M/1 QUEUING THEORY: AN OVERVIEW

'TAN YU TING. ^ZAITUL MARLIZAWATI ZAINUDDIN^BADRISYAH IDRIS

'Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Darul Ta'azim, Malaysia
^The UTM Centre for Industrial and Applied Mathematics (UTM-CIAM), 81310 UTM Skudai, Johor Darul Ta'zim, Malaysia
^Department of Neurosciences, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan

'sky yt(S<hotmail.com. *^zmarlizawati(S^utm.my. ^badrisvahfSkk.usm.mv

*Corresponding author

Abstract. Queuing is a common phenomenon in our daily life. Mathematical study on waiting line or queues is called queuing theory. Generally, queuing theory has been used extensively by service industry in order to optimize the service effectiveness and improve the customer satisfaction since it helps an organization to understand how a system operates while reviewing the efficiency of the system. Most of queuing theory deals with system performance in steady-state condition. That is, most queuing models assume that the system has been operating with the same arrival rate, service rate and other characteristics for a sufficiently long time that the probabilistic behavior of performance measures such as queue length is independent of initial condition. However, in many situations, the parameters defining the queuing system may vary over time. Under such circumstances, it is most unlikely that such systems are in equilibrium. This paper reviews the transient behavior (no assumption of statistical equilibrium) of the queuing model. The aim is to provide sufficient information to analysts who are interested in studying queuing theory with this special characteristic.

Afgyword? queuing theory; transient behavior

1.0 INTRODUCTION

Nowadays, many organizations face a challenging task to organize their processes more effectively and efficiently. Thus, understanding the system flow and analyzing system performance are critical as it takes important role in minimizing inefficiencies and delays. In general, an organization uses several forms of queue such as production lines and a wide variety of departments to tackle different situations and goals. At any given moment, there can be more people or cases requiring help or attention than a department can handle. The cooperation between the departments in accomplishing the tasks or delivering service leads to formation of queue. Hence, the efficiency of the company is highly related with queuing.

There is a mathematical study which describes the behavior of queue, that is queuing theory. Queuing theory involves the use of mathematical models to represent the behavior of customers for accessing a constrained resource. It helps an organization to understand how a system operates while reviewing the efficiency of the system. With the help of queuing theory, an organization can get a better understanding on how a system operates and improve its productivity.

Generally, most of queuing theory deals with system performance in steady-state condition. That is, most queuing models assume that the system has been operating with the same arrival rate, service rate and other characteristics for a sufficiently long time that the probabilistic behavior of performance measures such as queue length is independent of initial condition. However, most of the real queuing systems exhibit dynamic condition and time-dependent behavior. In many situations, the parameters defining the queuing system may vary over time. Under such circumstances, it is most unlikely that such systems are in equilibrium. The behavior of the system before the steady state is called transient behavior [1].

The investigation of the transient behaviour of the queuing system is imperative from the point of view of the theory and applications. By studying the transient behavior, we may investigate the effect to the system if there is a change in system parameter.

In this paper, an attempt is conducted to review the transient behaviour of queuing theory to identify the appreciate analysis technique according to the characteristic of queuing types. The remainder of this paper is organised as follows. Section 2 describes the component of queuing theory. Section 3 discusses the transient analysis of M/M/1 queue. Section 4 summarizes the conclusion.

2.0 COMPONENT OF QUEUEING THEORY

There are four terminologies to describe a queuing system, which are arrival process, service process, queue discipline and the way used by arrivals to join the queue. Normally, in arrival process, the arrivals are called customers. Meanwhile, in a service process, the servers are parallel if all the servers provide the same type of services and the customers only need to pass through one server to complete the service, but for service in series, the customers have to pass through several servers before the completing service. On the other hand, the queue discipline illustrates the way in which customers are served. According to Murthy [2], the components of the queuing system can be illustrated as Figure 1:

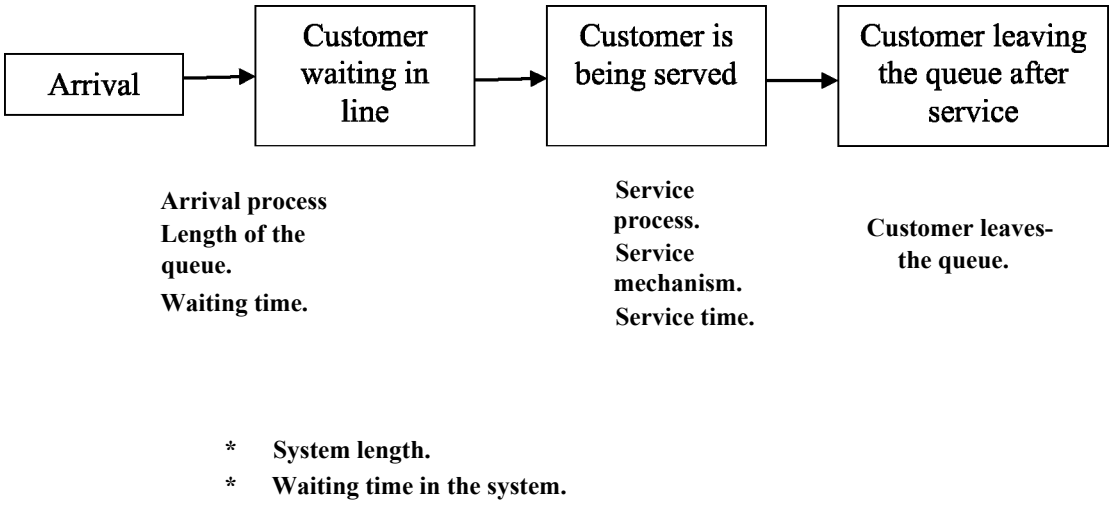

Figure 1: Components of Queuing System.

There are three key input parameters used in the queuing analysis, which are arrival rate, service rate and number of server [3]. These parameters play an important role to determine systems' queuing model and derive some useful performance measures such as the average waiting time, which is often adopted in optimization framework to improve the efficiency of a system.

When a queuing system begins operation, the state of the system will greatly affected by the initial state and by the time that has elapsed. After the sufficient time, the system becomes independent of the initial condition where the steady state condition exists. The probability distribution of the state of the system will remain the same over time under steady state condition.

In most of the queuing model, the parameters such as arrival rate, service rate and number of server are assumed to remain constant over time so that the steady state exists since the analysis of the transient condition is more complicated compared to the steady state condition. However, this assumption is seldom the case in real application. In a real queuing system, the arrival rate, service rate and the number of servers may vary over time. For example, a restaurant is likely to experience a larger arrival rate during afternoon than during other hours of the day. The number of server will vary during the day, with more servers available during busy periods.

When these parameters vary over time, the system is nonstationary and may never reach equilibrium. The probability distribution of the state of system also may vary over time. Thus, time-dependent performance characteristics are important in getting a deeper insight of a system's behaviour when short time intervals are of interest [4].

3.0 TRANSIENT ANALYSIS OF M/M/1 QUEUE

Most of the studies emphasized on steady state analysis of queuing theory since the transient analysis is more complex than the steady-state analysis. Due to these complexities, there are only few explicit expressions available even for simple model [5].

The M/M/1 queue is undoubtedly the simplest model for a queuing system. It is a single server queue characterized by Poisson process arrivals and an independent service time which is negative-exponentially distributed. It is relatively simple and yet the analysis of its transient behaviour leads to considerable difficulties [6]. Basically, the transient behaviors of single server queuing model have attracted the interest of researchers and treated with different methods since few decades ago.

In 1982, Upton and Tripathi [7] started to propose the approximation technique to study the transient behaviour of M(t)/M/1 queue where the arrival process can be described by discretely time varying exponential. In their opinion, the discrete time approach reduces the difficulties in analyzing the transient behaviour as the continuous time technique limit the computational flexibility and extensibility. A comparison is made between the proposed technique and imbedded Markov chain approach to conclude that the proposed method is more accurate and computationally efficient.

Using the same technique, Abate and Whitt [8] also investigated time-dependent behaviour of M/M/1 queue. They are concentrated more in obtaining the simple approximations that exposes the nature of transient behaviour. They highlighted that the factorial moment of the queue length as functions of time when the queue starts empty can facilitates developing approximation and the results are useful to demonstrate how the queue approaches steady state as time evolves.

Besides approximation technique, some researchers tried to explore the transient behaviour by employing generating function approach. The study by Van Den Berg and Groenendijkt [9] provided a transient analysis to estimate the distribution of the number of customer for M/M/1 queue with regularly changing arrival and service intensities, based on the generating function technique. A numerical example is conducted to show how the mean and variance of the number of the customers change over the period of time. Another study by Leguesdron ef a/. [10] also applied generating function approach to derive analytical formula for transient probability of M/M/1 queuing system.

Power series technique is another approach applied to seek the transient solution of queuing system. Sharma and Tarabia [11] considered a single server queue with finite waiting space. They developed a simple series form for the transient state probabilities of M/M/l/N queue using power series technique. Based on the proposed method, the state probabilities for steady state situations also can be derived straightaway.

Single server queuing system may deal with some special situations. For example, Abou-El-Ata e? a/. [12] discussed the transient solution of M/M/l/N with balking and reflecting barrier by using a simple and direct approach. By deriving the probability distribution of the number of customer, some performance measures such as expected number of customer in system and mean queue length are obtained. Moreover, they developed the state probabilities for steady state condition using same technique.

Generally, most of the queuing system assumed that there are no customers initially. However, in some situations, there is possibly that the customer exists in system at the beginning. Hence, Tarabia and El-Baz [13] further illustrated the studies of Sharma and Tarabia [11] to develop the transient solution for a nonempty **M/M/l/N** queue where there is an arbitrary number of initial customer in the system. Furthermore, Tarabia [14] also reviewed and derived a new series form for transient state probabilities for an **M/M/l/oo** queue with some customers presented in the system initially.

There are also some special queuing systems that the servers may not to start serving immediately although there are customers in system. Bohm and Mohanty [15] considered transient behaviour of M/M/1 queuing system with (M, N) policy in which the server starts serving only when there are N customers in queue and remain busy as long as there are at least M customers waiting. The transient solution is derived by using an alternative combinatorial method in which the queuing process is represented by a lattice path with diagonal steps. Further, they also analyzed the transient solution of an ordinary M/M/ 1 queue which formed by the assumption if there are only one customer in queue and no customer waiting in system.

Besides that, Sharda and Indra [16] concerned the single server queue with server on vacation providing service intermittently. That is a situation where the server may leave for some other important tasks after accomplishing service in hand and there is no customers waiting in queue. The transient mean queue length for some particular cases such as when the server is either busy or on vacation, when the server is either busy or available intermittently, and when the server is busy is developed by solving the different equation.

Another queuing system which usually occurs in our daily life is queue with repeated attempts. Queuing system with repeated attempts are characterized by the phenomenon that an arrival who finds the server are busy and will make a new attempt to get the service after a random time. Parthasarathy and Sudhesh [17] considered a single server retrial queue with Poisson arrival process and exponentially distributed service time in analyzing transient behaviour. By employing continued fraction method, they developed the time-dependent system size probabilities for state-dependent retrial queue. A numerical illustration to visualize the system size probabilities, expected system size and variance for the state-dependent rate, service and retrial rates are also provided to show how these solutions can be applied in practical situation.

Most of the studies emphasized on deriving the formula of transient behaviours, and seldom discuss about the application of transient solution in a real queuing system. Joy and Jones [18] attempted to study the transient probabilities for waiting list in a real case of hospital. They employed the lattice path to determine the transient probabilities of M*7M/1 queue with batch arrival in hospital. By plotting the performance curves for throughput, they investigated the impact of adding extra capacity into system as measured by an increasing service rate. The result showed that the chance to achieve the target of servicing 80% arrival in waiting list is high if extra capacity is introduced in the system. They also mentioned the importance of measuring the effect of the changes on the arrival rate and service rate in reducing the hospital waiting list.

The summary of the related works done on the transient analysis of M/M/1 queue discusses earlier is given in Table 1.

Author	Method	Queue Type
Upton and Tripathi [7]	Approximation approach	M(t)/M/1
Abate and Whitt [8]	Approximation approach	M/M/1
Van Den Berg and Groenendijkt [9]	Generating function technique	M/M/1 queue with regularly changing arrival and service intensities
Leguesdron ef a/. [10]	Uniformization and generating function technique	M/M/1
Sharma and Tarabia [11]	Power series technique	M/M/l/N
Abou-El-Ata e? a/. [12]	Direct approach	M/M/l/N with balking and reflecting barrier
Tarabia and El-Baz [13]	Power series technique	nonempty M/M/l/N
Tarabia [14]	Power series technique	nonempty M/M/l/oo
Bohm and Mohanty [15]	Combinatorial approach	M/M/1 with (M, N) policy
Sharda and Indra [16]	Generating function technique	M/M/1 with server on vacation providing service intermittently
Parthasarathy and Sudhesh [17]	Continued fraction	retrial M/M/1
Joy and Jones [18]	Lattice path approach	M*VM/1

Table 1: Classification of methods and queue types for transient solution.

4.0 CONCLUSIONS

In this paper, the transient behaviour of the M/M/1 queuing model has been reviewed. Sadly, the application of transient solution in a real case of queuing system is relatively less compare with steady state solution. The real queuing system generally exhibit dynamic condition where the arrival rate and service rate may change over time. Hence, employing the transient solution in real queuing system is more meaningful. It may help us to gain deeper insight on system performances as time evolves.

Understanding the system operation over the time is very important in capacity planning. Increasing servers' utilization and decreasing customers' waiting time can enhance system productivity. Queuing theory provides an effective and powerful modelling technique that can help managers in decision making for managing resources. Through an excellent capacity management, the efficiency and effectiveness of the available limited resources can be optimized and also improve the customer satisfaction.

It is suggested some characteristics of real-life queuing situations should be included for the purpose of future research work. For example, the characteristics of non-Poisson arrival process, non-exponential service distribution as well as the priority case should be incorporated in transient behavior of queuing system. Meanwhile, as an approach towards theoretical contribution, finding exact approximate numerical solutions for real-life queuing system are crucial. Such area is an important current research area as it is more practical by including specific characteristics or updating the model's prediction based on real word observations.

ACKNOWLEDGEMENT

This work is financed by MyBrainl5 provided by Ministry of Education Malaysia and we gratefully acknowledge the support.

REFERENCES

1. **Winston, W.L., Operaf:o?M 7?e.searc/:.' y4pp/:ca?:on and AJgon'fAm Fourth ed.** 2004**, United States: International Thomson.**
2. **Murthy, P.R., Operaf:o?M 7?asearcA. Second ed.** 2007**, New Delhi: New Age International (P) Ltd.**
3. **Koizumi, N., E. Kuno, and T. Smith, Afode&wg PaH'ewf 7%ws t/MHg a JVefwor% Health Care Management Science,** 2005**.** 8(1)**: p.** 49-60**.**
4. **Chydzinski, A.,** 7**ran.Henf o/ A/A/7V/G/7/A* gMene. Telecommunication Systems,** 2006**.** 32(4)**: p.** 247-262**.**
5. **Griffiths, J.D., G.M. Leonenko, and J.E. Williams, fran^H? ,so%Kf:oM ?c Operations Research Letters,** 2006**.** 34(3)**: p.** 349-354**.**
6. **Cahoy, D., F. Polito, and V. Phoha, Traym'eHf ReAav:cr Queues and /MafeJ ProceM&y. Methodology and Computing in Applied Probability,** 2013**: p.** 1-21**.**

7. **Upton, R.A. and S.K. Tripathi, An approjM'mafe fra?M:gwZ awa^ysfs q/^Ae M^j/M**/7 **^MCMg. Performance Evaluation,** 1982**.** 2(2)**: p.** 118-132**.**
8. **Abate,** J. **and W. Whitt, Transient AeAavior q/ zAe M/M// ^MgMg.* Aar^iwg a? ?Ae origin. Queueing Systems,** 1987**.** 2(1)**: p.** 41-65**.**
9. **Van Den Berg, J. and W. Groenendijkt,** 7**r<2/Mi'ewf Analysis o/* an M/M/**7 **QMeMe wi^A /^egM/ar/y Changing Arriva/ an^f Service** 7**n^ens:**7**ies. Teletraffic and DatatrafBc in a Period of Change. Proceedings oflTC,** 1991**.**13**.**
10. **Leguesdron, P., et al.,** 7**ra/MieMZ ana/ysis q/ M/M/**7 **^MeMe. Advances in Applied Probability,** 1993**: p.** 702-713**.**
11. **Sharma, O. and A. Tarabia, A simple fransienf ana/ysis q/aw M/M/**7**/N ^MeMe. Sankhya: The Indian Journal of Statistics, Series A,** 2000**: p.** 273-281**.**
12. **Abou-El-Ata, M.O., R.O. Al-Seedy, and K.A.M. Kotb, A &*ansienf so/M^ion q/ z/:e sfafe-M/M/**7**/N wifA &a%&ing anJ re/Zec^;'Mg Carrier. Microelectronics Reliability,** 1993**.** 33(5)**: p.** 681-688**.**
13. **Tarabia, A.M.K. and A.H. El-Baz, Exacf** 7**raw^'eMf ^o/Mfiowy q/ ^VowewpZy Markovian QMewes. Computers & Mathematics with Applications,** 2006**.** 52(6-7)**: p.** 985-996**.**
14. **Tarabia, A.M.K., A MewybrwnJaybr fAe Zramien? AeAavioMr q/" a MOH-eynpfy M/M/**7**/°° ^MgMe. Applied Mathematics and Computation,** 2002**.**132(1)**: p.** 1-10**.**
15. **Bohm, W. and S.G. Mohanty, r/:e frawMenf so/M^ion q/* M/M/**7 **^Mewes Mn^er (M,**7**V)-po/icy. A com^iwaforia/ approach. Journal of Statistical Planning and Inference,** 1993**.** 34(1)**: p.** 23-33**.**
16. **Sharda and Indra, Exp/ici? ^rawsien^ an<^ ^fea<**7**y s^a^e ZeHgfA ^ro^a^i/i^es q/ a ^MgMemg wMxM wifA server on vacafiow provi'<**7**mg service in^er**7**MiMen^y. Microelectronics Reliability,** 1995**.** 35(1)**: p.** 117-129**.**
17. **Parthasarathy, P.R. and R. Sudhesh,** 7**MMe-Je/?ef:<**7**eH? analysis q/ a single-server reZriaF ^MgMe wifA s?afe-JepeM<^eMf ra?es. Operations Research Letters,** 2007**.** 35(5)**: p.** 601-611**.**
18. **Joy, M. and S. Jones,** 7**raMs:eHf Pro&aM:?:esycr gKCMes wifA App^'caZiow^ fo Nospita/ Waifing Liyf Mawagemewy. Health Care Management Science,** 2005**.** 8(3)**: p.** 231-236**.**