International Journal of Grid and Distributed Computing
Vol. 11, No. 3 (2018), pp.11-22
http://dx.doi.org/10.14257/ijgdc.2018.11.3.02

Performance of M/M/1 and M/D/1 Queuing Models on Data Centers with Cloud Computing Technology Using MATLAB

N. Thirupathi Rao[1], PillaSrinivas[1] and K. Sudha[2], Debnath Bhattacharyya[1] and Tai-Hoon Kim[3*]

[1]Department of Computer Science & Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, AP, India
[2]Department of Computer Science & Engineering
Lenora College of Engineering, AP, India
[3] Sungshin Women's University, Bomun-ro 34da-gil, Seongbuk-gu, Seoul, Korea
[1]nakkathiru@gmail.com,[1]srinivasp3@gmail.com,[1]debnathb@gmail.com,
[3]taihoonn@daum.net*

Abstract

Cloud computing was the technology developed to store the info and support the users with the access to the info hold on by charging a token quantity for the storage of information and for providing necessary steps for storing the info and fro providing security to the info that was hold on. The content hold on in varied servers at varied locations supported the sort and size of the content. The content may be accessed to the users with valid registrations and a group of security verifications entered by the purchasers. The content that was hosted within the servers can even be used for hosting varied applications and varied alternative set of choices of systems in varied fields. It's one among the foremost illustrious and principally used analysis areas within the recent years for additional development in varied set of applications and its usages associated with many set of shoppers within the real time setting. Performance analysis in cloud computing has been another major thrust space within the recent past, that is of crucial interest for each cloud suppliers and cloud customers. Solely few notable works are revealed with regards to performance analysis in cloud computing. Typically analytical models established for assessing the operating and therefore the performance of cloud server farms may be studied beneath kind of configurations and assumptions are supported queuing theory and its accuracy is verified with numerical calculations and simulations. The issues at hand create to the task of evaluating the performance of information center with varied queuing models to grasp the distribution of the performance parameters with arrival and repair rates, traffic intensity, range of servers and therefore the associated possibilities. The goals of this thesis is to supply a framework through programs associated with queuing models and value the performance parameters, try validation, sensitivity analysis and build comparisons for information centers. Gift thesis evaluates the performance parameters of cloud information centers supported queuing theory for each single server and multi-server models. The steady state performance parameter formulations known are programmed in MATLAB® setting. The models considered for evaluation for single servers include M/M/1, M/G/1, M/D/1. Service rates have a wider range of distributions including exponential, generalize and Erlang type.

***Keywords**: Cloud computing, queuing models, exponential distribution*

Received (December 26, 2017), Review Result (February 16, 2018), Accepted (February 18, 2018)

[*] Corresponding Author

ISSN: 2005-4262 IJGDC

1. Introduction

The cloud is wherever one will use technology once required, as long together wants it. The clouds are often each code and infrastructure service. In terms of maturity, code is far additional evolved than hardware within the cloud. The clouds are often associate degree application one will access through the net or a server. With cloud computing, the foremost advantage for users was that the programs that may be associated with code programs associated with many applications hold on within the machine couldn't be dead within the native computers or laptops. All the programs associated with the applications that were being connected to the cloud mechanism were being accessed by the web to figure. The web association was shall within the field of cloud processing and its connected application areas. The foremost advantage for the purchasers and therefore the users was the crash report. Whenever a crash happens within the system or the system connected to the cloud, the information are often still accessible because the actual data was hold on within the servers of the cloud not within the actual systems that were being connected to the cloud setting. The information won't be disturbed or any harm to the particular data because it was hold on at varied servers connected one another and settled at varied locations. The users may be customers from varied various corporations, various servers and therefore their connected applications and therefore the various networks that can be accustomed connect of these servers and the machines. Here, the information was hold on in varied servers with varied configurations and varied operational systems and everyone these servers were settled at varied locations, so the information are often secured despite the fact that a significant harm or loss happens for servers settled at one place on the country or on earth.

1.1. Cloud Data Centers

Cloud computing was the technology and model that was aimed to supply the services to varied set of consumers by the model of paying some certain quantity for utilizing sure set of operations and tasks from the system model. By exploitation this model, the users will use the resources from numerous set of operations like networks, servers, applications and numerous services associated with the system. The user will utilize the services and to transfer the info or the programs or the set of codes that were keep in numerous servers and set at numerous locations. The info may be downloaded or the utilized by locating at numerous places by merely having a certified access to the set of consumers. The shoppers were supported the set of applications they were doubtless to use or they need to use the opposite set of information within the returning future conjointly. The trouble unbroken by users for providing knowledge to the servers, maintaining the info within the servers and providing security to the info servers were terribly minimum. Most of the tasks and also the security step for providing the system and also the machines and servers are going to be taken care by the service suppliers and also the individuals whoever maintaining the services.

Data centers were being maintained at numerous locations of the planet at numerous countries within the continents. The users associated with that individual center will ready to transfer the info at any purpose of your time. The servers and also the machines at every knowledge center were internally connected to every different such to keep up the info unambiguously and to supply best service to the shoppers. The information that was keep at numerous data centers were being shared by numerous analysis organizations, remote process applications and different connected applications. Some organizations have to be compelled to be used principally for numerous applications and also the knowledge ought to be secured such that a lot of security ought to be provided to the knowledge that was being keep by numerous analysis organizations and different international corporations UN agency can have a large set of transactions and also the customers connected data.

Figure 1. Cloud Data Center

1.2. Queuing Systems

The queuing systems that were within the type of theoretical model were meant to develop and supply the varied set of models for predicting and estimating the performances of assorted systems subject to the random in nature of the systems. The history of the queuing theory was 1st started within the years of 1908. The analysis of the waiting lines may be analyzed simply with the assistance of queuing theory models. It can also be treated because the analysis of the waiting lines or queues at numerous places may be simply analyzed with the assistance of those queuing theory models. These models area unit extremely renowned and helpful in predicting the performance of the varied systems virtually precisely the same condition of their actual behaviors. The analysis on numerous operations during a system referred to as the research is additionally one in all the elements of the queuing systems or queuing theory models. The future average values for a system may be expected or may be analyzed by these queuing systems. The input file that we have a tendency to area unit reaching to provide to a systems or a machine was measured for associate extended amount of your time. The queuing systems can assume that the arrival times and repair times area unit continuously random in nature. The following figure shows the queuing system model. The contents of the system area unit the client arrivals *i.e.*, the amount of users or the shoppers UN agency were victimization or going in the queue for obtaining the service or to examine one thing or get set of taking one thing from a degree. once the shoppers enters, if the road is free they'll simply get their service, if the road United States busy then sure time are taken for every user to induce their application or their task are completed.

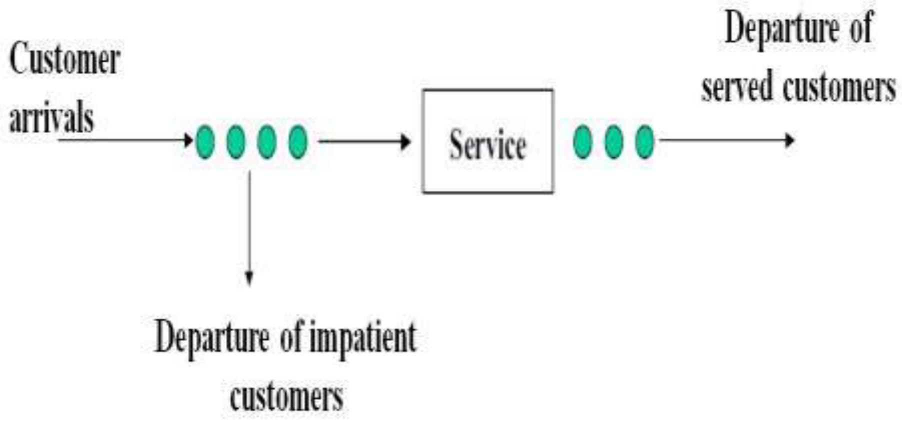

Customer arrivals
Service
Departure of served customers
Departure of impatient customers

Figure 2. Queuing System

Various set of assumptions were involved in queuing systems whenever we are trying to solve the issues and the problems related to the queuing systems and its related applications. The arrivals to a queuing system or the system that was being solved by using the queuing system model are independent distribution, exponential distribution. The queues which were in heavy lines or small lines will not discourage the customers.

2. Problem Description and Solution

The problems at hand bring about to the task of evaluating the performance of information center with numerous queuing models to know the distribution of the performance parameters with arrival and repair rates, traffic intensity, range of servers and therefore the associated possibilities. The goals of this thesis is to supply a framework through programs associated with queuing models and evaluates the performance parameters, try validation, sensitivity analysis and create comparisons for information centers. Gift thesis evaluates the performance parameters of cloud information centers supported queuing theory for each single server and multi-server models. The steady state performance parameter formulations known are programmed in MATLAB® surroundings. The models thought of for analysis for single servers embody M/M/1, M/G/1, M/D/1 and M/Er/1. The multi-server models thought of are M/M/c, M/M/c/c, M/M/c/K and M/M/c/c+r. Interarrival rates for all the higher than models have exponential distribution. Service rates have a wider vary of distributions as well as exponential, generalized, and settled and telephone unit kind.

Major drawback in cloud computing is knowing the character of cloud server performance, that rely on the analysis and utilization of optimum performance parameters. Therefore there's a necessity to analytically model the cloud servers/data centers mistreatment queuing systems and estimate the performance parameters. Usually numerous analytical models designed and developed for analyzing the performance of assorted cloud server applications with various set of assumptions and configurations. These assumptions are supported queuing theory and its accuracy is often verified with numerical calculations and simulations.

2.1. Proposed System

Present system deals with the performance evaluation in-terms of steady state parameters of a small cloud server farm using single and multi server queuing models. Single server models include *M/M/1*, *M/G/1*, *M/D/1* and *M/Er/1*. Multi-server model considered include *M/M/c*, *M/M/c/c*, *M/M/c/K* and *M/M/c/c+r*. A comparison among the steady state parameters evaluated for the above queuing models with respect to traffic intensity along with sensitivity analysis is also proposed.

3. System Design

System design includes parameters, performance measures, stability and properties considered for the data center performance evaluation using queuing models. System parameters are,

a. It is customary to introduce some notation for the performance measures of interest in queuing systems.

b. ***Number of customers in the system (L_S)***: In steady-state, the expected value of the state distribution gives the mean number of customers in the system.

c. ***Number of customers in the queue (L_Q)***: In steady-state, the expected value of the state distribution in a queue gives the mean number of customers in the queue.

d. ***Utilization (or) Traffic intensity* (ρ):** For a queuing system with a single server, utilization ρ is the fraction of time the server is busy. When there is no limit on the capacity of the system, then

ρ = mean arrival rate/mean service rate =λ/μ

e. The utilization when there are multiple servers (c), is the mean fraction of busy servers. Since ***c***is the overall service rate, in this case ρ = . For a stable (ergodic) system, the condition for stability is ρ< 1.

f. ***Throughput ( )***: The throughput for a queuing system with infinite capacity is the mean number of customers processed in a unit of time, *i.e.* the departure rate. Since the departure rate is equal to the arrival rate (and assuming ρ< 1), the throughput is =c ρ**.** For a queuing system with finite capacity, there can be loss in the systems, and so the throughput can be less than the arrival rate. In this case, throughput is often denoted differently (*e.g.* as *S*) **to distinguish it from the arrival rate.**

g. ***Response Time* (W_S):** (or sojourn time) It is the total time a customer spends in the system.

h. ***Waiting Time* (W_Q):** It is the time a job spends in the queue waiting to be serviced. Therefore, response time is the sum of the waiting time (W_Q) and the service time(***1/)*** for a customer *i.e.* $W_S = W_Q+(1/\mu)$

To evaluate the performance parameters of data center in cloud architecture, the programs consider the following input values:

i. Interarrival rates ,

ii. Service rates ,

iii. Number of servers , *c*

iv. Maximum number of customers allowed , *K*

v. Waiting capacity of customers , *r*

vi. Coefficient of variance , C_{oV}

vii. Erlang parameter , *Er*

The list of output parameters of the programs are highlighted below:

i. Length of customers in a system , L_S

ii. Length of customers in queue , L_Q

iii. Waiting time of customers in a system , W_S

iv. Waiting time of customers in queue , W_Q

v. Associated probabilities

4. System Implementation

Whenever if there is any uncertainty in arrival and service times of any system or application, the queuing models are the best fitted application or the model which could be used to estimate the presentation of the systems for various services to the users or customers. For the present work it is assumed that customers are served in the order in which they arrive in the system (First-Come-First-Served or FCFS). The MATLAB programs and formulations are given at Appendix 1.

4.1. Queuing Models

a) Queuing Model - M/M/1

The M/M/1 queue model has Interarrival times, which are exponentially distributed with parameter and also service times with exponential distribution with parameter. The system has only a single server (c=1) and uses the FIFO service discipline. The exponential distribution has a squared coefficient of variation of 1. The waiting line is of infinite size. The M/M/1 system is a pure birth-/death system, where at any point in time at most one event occurs, with an event either being the arrival of a new customer or the completion of a customer's service. What makes the M/M/1 system really simple is that the arrival rate and the service rate are not state dependent.

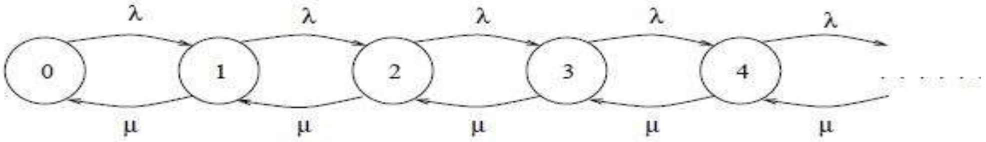


Figure 3. Markov Chain for M/M/1 Queue

Steady state performance measures for the above model are:

$$L_S = \frac{\rho}{1-\rho} \quad ; \quad L_Q = \frac{\rho^2}{1-\rho}$$

$$W_S = \frac{1}{\mu(1-\rho)} \quad ; \quad W_Q = \frac{\rho}{\mu(1-\rho)}$$

$$P\ (W_S < t) = 1 - e^{-(1-\rho)\mu t} \quad ; \quad P\ (W_S > t) = e^{-(1-\rho)\mu t}$$

$$P\ (W_Q <= t) = 1 - \rho e^{-(1-\rho)\mu t} \quad ; \quad P\ (W_Q >= t) = \rho e^{-(1-\rho)\mu t}$$

Note: P is the notation for probability

b). Queuing Model - M/D/1

The M/D/1 queue has Interarrival times, which are exponentially distributed with parameter and also service times with constant distribution with parameter. The system has only a single server (c=1) and uses the FIFO service discipline. The waiting line is of infinite size. Deterministic distribution (with constant service times) has a zero variance for this distribution. For this reason one can achieve always the highest throughput (lowest delays) for deterministic service times. The simplest important result is that the average number waiting is half that waiting with exponentially distributed service.

Steady state performance measures for the above model are:

$$L_S = \rho + \frac{\rho^2}{2(1-\rho)} \quad ; \quad L_Q = \frac{\rho^2}{2(1-\rho)}$$

$$W_S = \frac{2-\rho}{2\mu(1-\rho)} \quad ; \quad W_Q = \frac{\rho}{2\mu(1-\rho)}$$

$$Var\,(Ls) = \frac{\rho^3}{3(1-\rho)} + \frac{\rho^2}{2(1-\rho)} + 2\rho^2(3-2\rho)(1-\rho) + \rho(1-\rho)$$

$$Var\,(Ws) = 2Wq^2 + \frac{2\rho}{\mu^2 3(1-\rho)} + \frac{1}{\mu^2(1-\rho)} - Ws^2$$

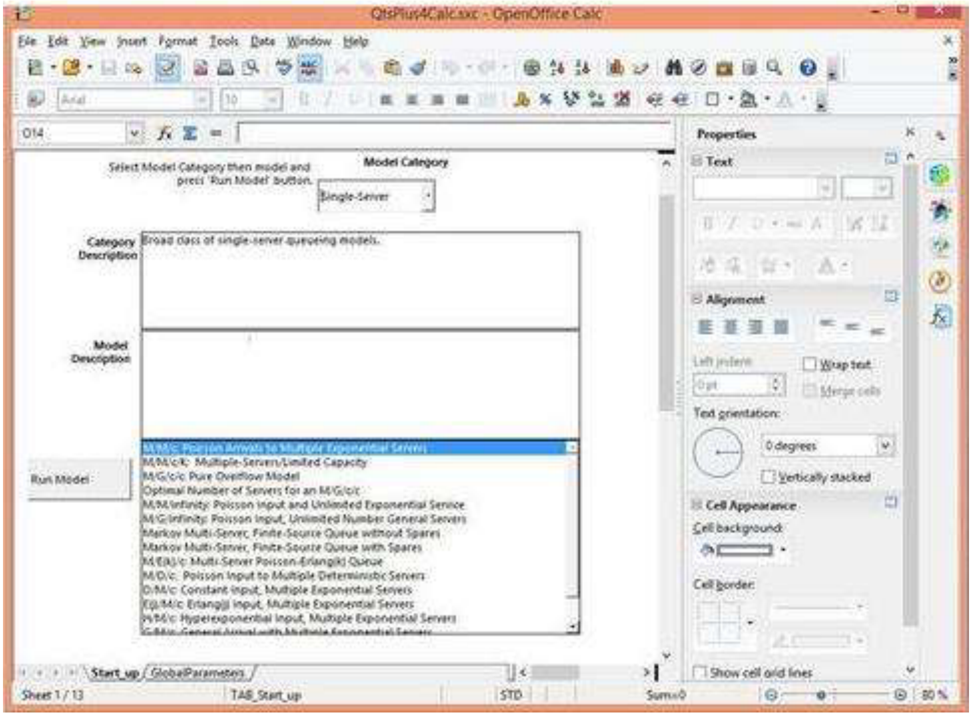
$$Var\,(Lq) = \frac{\rho^3}{3(1-\rho)} + \frac{\rho^2}{2(1-\rho)} + 2\rho^2(1-\rho)$$

$$Var\,(Wq) = Wq^2 + \frac{2\rho}{\mu^2 3(1-\rho)}$$

c). QtsPlus4Calc Software (*http://qtsplus4calc.sourceforge.net*):

QtsPlus4Calc, release 2006 is a freeware developed by Donald gross and Carl M. Harris of George Mason University. This software provides a platform to evaluate performance of various queuing models. A sample screenshot of the software environment is given at Figure 12. The calculator includes single server, multi-server, priority, bulk and network models.

Numerical Solutions to Queuing Systems *(http://queueing-systems.ens-lyon.fr):*Numerical solutions to queuing systems software provide a platform to evaluate performance of various queuing models with generalized service distributions. A sample screenshot of the software environment is given at Figure 3. The calculator includes single server models.



Select Model Category then model and press 'Run Model' button.

Model Category: Single-Server

Category Description: Broad class of single-server queueing models.

Model Description

Run Model

- M/M/c: Poisson Arrivals to Multiple Exponential Servers
- M/M/c/K: Multiple-Servers/Limited Capacity
- M/G/c/c Pure Overflow Model
- Optimal Number of Servers for an M/G/c/c
- M/M/Infinity: Poisson Input and Unlimited Exponential Service
- M/G/Infinity: Poisson Input, Unlimited Number General Servers
- Markov Multi-Server, Finite-Source Queue without Spares
- Markov Multi-Server, Finite-Source Queue with Spares
- M/E(k)/c: Multi-Server Poisson-Erlang(k) Queue
- M/D/c: Poisson Input to Multiple Deterministic Servers
- D/M/c: Constant Input, Multiple Exponential Servers
- E(j)/M/c: Erlang(j) Input, Multiple Exponential Servers
- H/M/c: Hyperexponential Input, Multiple Exponential Servers
- G/M/c: General Arrival with Multiple Exponential Servers

Figure 4. QtsPlus4Calc Environment Screenshot

6. Results and Discussion

Based on the queuing models discussed in the previous sections, input parameters limits were identified from literature for performance evaluation of small cloud computing data center farm (*i.e.* number of servers, *c* were limited to 4 and 8).The inter arrival and service rates are chosen for a range of traffic intensity (or utilization) varying from 1 - 2. Each time service rate was varied to values 3-4 and the respective performance was evaluated. Input parameter limits are given at Table.3

Table 3. Input Parameter Limits

Parameter	Chosen Limits
μ	0- 2
λ	3-4
ρ	0- 1
c	4, 8
K	4, 8
r	4, 8
^{c}ov	1.5
Er	2, 4
t	1- 2

Evaluation of Cloud Computing Data Center Performance with M/M/1 Queuing Model:

Table 4. Performance of Data Center - M/M/1, Model = 1

μ	**ρ**	*Ls*	*Lq*	*Ws*	*Wq*	*t*	*P(Ws <t)*	*P(Ws >t)*	*P(Wq <=t)*	*P(Wq> =t)*
3	0	0.00	0.00	1.00	0.00	1	0.00	1.00	1.00	0.00
3.1	1.1	0.11	0.01	1.11	0.11	1.1	0.07	0.81	0.92	0.06
3.2	1.2	0.25	0.05	1.25	0.25	1.2	0.18	0.74	0.87	0.21
3.3	1.3	0.43	0.13	1.43	0.43	1.3	0.22	0.63	0.79	0.38
3.4	1.4	0.67	0.27	1.67	0.67	1.4	0.28	0.58	0.64	0.45
3.5	1.5	1.00	0.50	2.00	1.00	1.5	0.34	0.65	0.55	0.59
3.6	1.6	1.50	0.90	2.50	1.50	1.6	0.22	0.71	0.49	0.67
3.7	1.7	2.33	1.63	3.33	2.33	1.7	0.14	0.84	0.37	0.79
3.8	1.8	4.00	3.20	5.00	4.00	1.8	0.08	0.89	0.28	0.84
3.9	1.9	9.00	8.10	10.00	9.00	1.9	0.06	0.94	0.14	0.92
4	2	Inf	Inf	Inf	Inf	2	0.00	1.00	0.00	1.00

Table 5. Performance of Data Center - M/M/1 Model= 2

μ	ρ	*Ls*	*Lq*	*Ws*	*Wq*	*t*	*P(Ws<t)*	*P(Ws>t)*	*P(Wq<=t)*	*P(Wq>=t)*
3	0	0.00	0.00	0.50	0.00	0	0.00	1.00	1.00	0.00
3.1	1.1	0.11	0.01	0.56	0.06	1.1	0.06	0.92	0.96	0.06
3.2	1.2	0.25	0.05	0.63	0.13	1.2	0.19	0.81	0.81	0.21
3.3	1.3	0.43	0.13	0.71	0.21	1.3	0.28	0.70	0.76	0.38
3.4	1.4	0.67	0.27	0.83	0.33	1.4	0.32	0.59	0.69	0.45
3.5	1.5	1.00	0.50	1.00	0.50	1.5	0.45	0.51	0.59	0.59
3.6	1.6	1.50	0.90	1.25	0.75	1.6	0.32	0.59	0.42	0.67
3.7	1.7	2.33	1.63	1.67	1.17	1.7	0.28	0.71	0.38	0.79
3.8	1.8	4.00	3.20	2.50	2.00	1.8	0.19	0.82	0.27	0.84
3.9	1.9	9.00	8.10	5.00	4.50	1.9	0.06	0.93	0.13	0.92
4	2	Inf	Inf	Inf	Inf	2	0.00	1.00	0.00	1.00

Evaluation of Cloud Computing Data Center Performance with M/D/1 Queuing Model,

Table 6. Performance of Data Center - M/D/1 Model= 1

Variance										
μ	**λ**	***ρ***	***Ls***	***Lq***	***Ws***	***Wq***	***Var(Ls)***	***Var(Ws)***	***Var(Lq)***	***Var(Wq)***
1	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1	0.1	0.1	0.11	0.01	1.06	0.06	0.14	0.08	0.02	0.08
1	0.2	0.2	0.23	0.03	1.13	0.13	0.33	0.18	0.07	0.18
1	0.3	0.3	0.36	0.06	1.21	0.21	0.53	0.33	0.14	0.33
1	0.4	0.4	0.53	0.13	1.33	0.33	0.72	0.56	0.25	0.56
1	0.5	0.5	0.75	0.25	1.50	0.50	0.90	0.92	0.40	0.92
1	0.6	0.6	1.05	0.45	1.75	0.75	1.14	1.56	0.67	1.56
1	0.7	0.7	1.52	0.82	2.17	1.17	1.73	2.92	1.34	2.92
1	0.8	0.8	2.40	1.60	3.00	2.00	3.93	6.67	3.67	6.67
1	0.9	0.9	4.95	4.05	5.50	4.50	19.12	26.25	18.99	26.25
1	1	1	Inf	Inf	Inf	Inf	Inf	NaN	Inf	Inf

Table 7. Performance of Data Center - M/D/1 Model= 2

Variance										
μ	λ	ρ	Ls	Lq	Ws	Wq	Var(Ls)	Var(Ws)	Var(Lq)	Var(Wq)
2	0	0	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00
2	0.2	0.1	0.11	0.01	0.53	0.03	0.14	0.02	0.02	0.02
2	0.4	0.2	0.23	0.03	0.56	0.06	0.33	0.05	0.07	0.05
2	0.6	0.3	0.36	0.06	0.61	0.11	0.53	0.08	0.14	0.08
2	0.8	0.4	0.53	0.13	0.67	0.17	0.72	0.14	0.25	0.14
2	1	0.5	0.75	0.25	0.75	0.25	0.90	0.23	0.40	0.23
2	1.2	0.6	1.05	0.45	0.88	0.38	1.14	0.39	0.67	0.39
2	1.4	0.7	1.52	0.82	1.08	0.58	1.73	0.73	1.34	0.73
2	1.6	0.8	2.40	1.60	1.50	1.00	3.93	1.67	3.67	1.67
2	1.8	0.9	4.95	4.05	2.75	2.25	19.12	6.56	18.99	6.56
2	2	1	Inf	Inf	Inf	Inf	Inf	NaN	Inf	Inf

7. Conclusions

Performance analysis for single servers indicate that because the service rate (r) will increase for a relentless vary of traffic intensity (ρ) solely waiting times of consumers within the system (WS) and queue (WQ) decreases, wherever because the length of consumers in system (LS) and queue (LQ) stay unchanged because it is freelance. For identical input parameters M/D/1 model shows optimum performance in terms of queue lengths and waiting times followed by M/Er/1, M/M/1. Performance of M/D/1 shows prejudicial nature compared with different queuing models that are attributed to higher price of CoV. For higher order telephone unit parameters, M/G/1 and M/Er/1 models behave in shut comparison.

Multiple server (c = 3,4) performance analysis indicates that because the service rate will increase for a relentless vary of traffic intensity (ρ), similar nature as within the case of single servers is determined with regards to queue lengths and waiting times. Performance analysis of tiny cloud computing information center is mentioned with the idea supported queuing systems. Single server and multiple server models square measure bestowed at the side of their formulations for performance parameters. MATLAB programming/code generation and implementation for performance analysis of cloud computing information server frame is accomplished. Comparisons among varied models square measure tried and relevant observations square measure highlighted.

References

[1] J. S. Chandrakala, “Survey on Models to Investigate Data Center Performance and QoS in Cloud Computing Infrastructure”, First International Conference on Recent Advances in Science & Engineering, **(2014)**.

[2] M. Hlynka, “Comparing expected wait times of a M/M/1queue”, Department of Mathematics and Statistics, University of Winsor, **(2010)**.

[3] H. Khazaei, Jelena and Vojislav, “Performance Analysis of Cloud Computing Centers Using M/G/m/m+r Queuing Systems”, IEEE Transactions on parallel and distributed systems, vol.23, **(2012)**.

[4] J. Sztrik, “Basic Queuing Theory”, University of Debrecen, Faculty of Informatics, **(2012)**.

[5] N. Khanghahi and R. Ravanmehr, “Cloud Computing Performance Evaluation: Issues and Challenges”, International Journal on Cloud Computing Services and Architecture, vol.3, **(2013)**.

[6] T. V. Mathew, "Queuing Analysis", Transportation Systems Engineering, Indian Institute of Technology, Bombay, **(2014)**.

[7] H. Khazaei, "Performance Modeling of Cloud Computing Centers", Doctoral dissertation, The University of Manitoba, Canada, **(2012)**.

[8] B. Yang, F. Tan, Y. Dai and S. Guo, "Performance evaluation of cloud service considering fault recovery", First International Conference on Cloud Computing (CloudCom) 2009, **(2009)**.

[9] I. Adan and J. Resing, "Queuing Systems", Eindhoven University of Technology, The Netherlands, **(2015)**.

[10] T. Sai Sowjanya, D. Praveen, K. Satish and A. Rahiman, "The Queuing Theory in Cloud Computing to Reduce the Waiting Time", IJCSET, vol.1, **(2011)**.

[11] A. Brandwajn and H. Wang, "A conditional probability approach to M/G/1 – like queues", Performance Evaluation, vol.65, **(2008)**.