Cluster Computing
https://doi.org/10.1007/s10586-017-1640-7

An efficient performance evaluation model for the resource clusters in cloud environment using continuous time Markov chain and Poisson process

M. Kowsigan[1] · P. Balasubramanie[2]

Received: 31 October 2017 / Revised: 20 December 2017 / Accepted: 26 December 2017

Abstract
Cloud computing enables us to share the jobs across various cloud data centers with the internet as a backbone in order to process the jobs. The selections of the data centers to process the jobs are based on the potential of the data centers, arrival rate of the jobs and better resource utilization. Performance modeling and evaluation of the cloud systems shows a perfect picture of probability of distribution of jobs among the datacenters, mean number of jobs in the datacenters and makespan of the cloud environment. In order to provide a best performance evaluation model which leads to a prosperous cloud environment the proposed approach calculates the arrival rate of jobs using Poisson process in which the arrival rate can be calculated for infinite time intervals and the mean number of jobs and also the resource utilization are modeled using continuous Markov chain. The problems related to the calculation of arrival rate and resource utilization model of the existing approaches are solved to enhance the job scheduling by minimizing the makespan of the cloud environment.

Keywords Cloud computing · Response time · Resource utilization · Performance evaluation · Poisson process · Markov Chain

1 Introduction

Cloud computing allows the provider to issue the computing resources such as storage technologies, software and platforms as a common service to the users in a rental basis when it is on-demand. The concept of virtualization plays a vital role in the base of the cloud computing which enables to form the groups of computing resources from the server clusters and allocates or reallocates the computing resources to the application in an on demand manner [1]. One of the major challenge in cloud computing is to minimize the makespan of the cloud environment by efficiently scheduling the available resources to the incoming jobs. Some of the parameters such as arrival rate of the jobs, response time and waiting time of the job at the resource clusters are considered while scheduling the jobs to the available resources in a cloud environment. In this paper some of the difficulties faced by the existing approaches related to the arrival rate and device utilizations are addressed and solved.

The execution time or completion time of jobs at a particular cloud environment is called makespan of that cloud environment [2]. Resources are clustered based on their potentials which lead to an efficient scheduling in a cloud environment. In the proposed approach resources are clustered by calculating the potential of each and every resource by using the equivalence partitioning recurrent node weight algorithm [3]. The main objective of the proposed approach is to develop a suitable performance and evaluation model for the parameters such as arrival rate, completion time and resource utilization of a cloud environment. The service time of the jobs mainly depends on the arrival rate of the jobs at the computing resources. The arrival rate is calculated by Poisson process in the proposed approach due to two reasons: one is Poisson process obeys the properties of the Markov chain which is systematically manageable and second one is arrival rate of jobs can be calculated at inde-

✉ M. Kowsigan
kowsigan_m@rediffmail.com

P. Balasubramanie
balasubramanie@rediffmail.com

1 Department of Information Technology, Sri Krishna College of Technology, Anna University, Coimbatore, India

2 Department of Computer Science & Engineering, Kongu Engineering College, Anna University, Erode, India

Published online: 19 February 2018

pendent time intervals and also no limit for the time intervals [4].

Resource Utilization can be calculated by identifying the mean number of jobs residing in the resource clusters. Markov chain model is used to identify the probability of mean number of jobs resides in the resource clusters by ensuring that execution times at the resource cluster are distributed exponentially and the continuous arrival rate of the jobs are also exponentially distributed. The Markov chain is divided into two types as discrete time Markov chain and continuous time Markov chain. The occurrence of the events are only at the end of the time step in discrete time Markov chain whereas in continuous time Markov chain the occurrence of events depends on independent time intervals [5]. Continuous time Markov chain is used in the proposed approach to identify the probability of resource utilization by calculating the arrival rate of the jobs at independent time interval using Poisson process. To produce an efficient performance evaluation model this paper overcomes some of the disadvantages in the existing approaches and produces the near optimal solution.

2 Related works

Many literatures are available for performance evaluation model regarding to the job scheduling on a cloud environment. To achieve high efficiency, a mathematical model for performance evaluation of the cloud resource farms was implemented to calculate the approximate distribution of jobs and mean number of jobs in the system [6]. The multilevel feedback queue scheduling was examined by Markov chain model under some common scheduling policies to determine the CPU processing time [7]. Markov chain analysis had been used to evaluate the Drum Buffer Rope with the other conventional approaches to show the improved performance of the Drum Buffer Rope in the Job shop environment [8]. Multi site partitioning problem in mobile clouds had been addressed and solved using energy efficient multisite offloading scheme and analyzed by Markov decision process in order to achieve the minimum energy consumption of mobiles [9].

Photovoltaic power of grids was represented by a semi Markov process to produce a reliable Photovoltaic power and also schedules the energy storage units in stochastic manner in concurrence with other resources [10]. The simulation results of canonical LARAC algorithm justifies that one climb offloading scheme was well suited for Markovian probabilistic channel and reduces the energy consumed by the mobile devices [11]. In mobile clouds the fault tolerance can be achieved by declining the volatility property of mobile devices and prediction of prediction of states with the help of Markov chain model [12]. An efficient scheduling of virtual machines in cloud can also be attained by predicting the future bandwidth of all the virtual machines with the help of accumulating the prior knowledge of the bandwidth workload which leads to vigorous modification in Hidden Markov Model in order to manage the telehealth system [13].

One of the important factors for simulations and performance evaluation is workload generation. A workload generator called BURSE was proposed for both the characteristics of workloads bursty and self similar in cloud computing based on the nature and behavior of the matter of two states Markov Modulated Poisson processes [14]. The two important parameters of mobile cloud offloading are the utilization of energy by the mobile extreme and the response time incident by the job of the mobile device. The parameters such as energy response time weighted sum, energy response time product and the energy response time weighted product are examined and applied to the different offloading strategies of mobile cloud to conclude the performance evaluation model in an effective manner [15]. The job scheduling and potential of the bystander workloads of the cross virtual machine covert channel was evaluated using Continuous Markov model. An efficient virtual machine allocation was used to minimize the error rate in cross virtual machine covert channel [16].

The concept of reinforcement learning was used for task scheduling and the concept of queuing theory was used to optimize the task scheduling based on the resource constraints and it also results in disclose of relationship between the arrival rate, resource utilization and storage size [17]. The resources are allocated effectively by predicting the usage of virtual machine CPU as well as considering both long and short term behaviors of VM's by implementing Hurst exponent which was an companion of Markov transition model [18]. Markovian decision process was used as a framework to achieve the near optimal solution for the dynamic offloading problem in mobile cloud computing [19]. Markov chain prediction technique was used by nodes scheduling approach to analyze the big streaming data and ensures that big streaming data can be effectively processed [20]. The methodologies of the above literatures are not perfectly suited for the performance evaluation of the resource clusters in cloud computing in the case of infinite arrival rate and large number of resource clusters. By generally speaking in such situations undergo the some of the probability distribution models as Poisson process for calculating the arrival rate and Markov chain for estimating the resource utilization as well as mean number of jobs in the resource cluster. The hidden objective of this work specifies the effective job scheduling in cloud environment. An improved metaheuristic approach was used to enhance the job scheduling in cloud environment [21]. Probability distribution method had been used to improve the job scheduling in cloud environment [22].

3 Proposed methodology

The cloud environment contains many servers which schedules 'n' number of jobs to the available resource clusters. Each server maintains an order among the jobs for its own scheduling and selects the best resource cluster by using equivalence partitioning recurrent node weight algorithm [1] to perform the operation on its jobs. In equivalence partitioning recurrent node weight algorithm, the potential of the resource clusters are calculated based on the three steps. First step is to assign the weight to the nodes and make an identity to those nodes. Second step is to establish the connection between the nodes in the environment for an effective communication among them. Third step is to separate the resource king from other resources as mentioned by the procedure of the algorithm. Makespan is an important parameter in a cloud environment while scheduling the jobs. Makespan is defined as the overall completion time of jobs for a particular cloud environment and it mainly focuses on the arrival time of the job, response time of the jobs as well as the device utilization. The makespan of a cloud environment is calculated as the time between the jobs enters into the resource cluster and exits the resource cluster inclusive of jobs entering into the same resource cluster as well as different resource cluster several times. All the resource clusters are interconnected among themselves through a network medium to ensure the neighborhood facilities among the clusters. The servers use the sequential ordering to serve the jobs. Every resource cluster in the cloud environment accepts jobs either internally or externally. In the existing approaches the selection of resource clusters was based on probabilistic approach which leads to an inefficient solution. In order to obtain near optimal solution, the best resource clusters are selected by calculating the resource potential using equivalence partitioning recurrence node weight algorithm. The jobs on resource cluster A is carried to resource cluster B with the cluster potential CP_{AB} or departures from the resource cluster if the potential of resource cluster satisfies the jobs with potential $CP_{A,out} = 1 - \sum_B CP_{AB.}$

The overall arrival rate into the resource cluster A is the sum of rate of jobs arriving to resource cluster A externally and rate of jobs arriving to resource cluster A internally in the cloud environment.

Overall arrival (λ_A) = External arrival $(e_A) + \sum_B \lambda_B\, CP_{BA}$

At the same time all the parameters of the equations are solved to identify the overall λ_A's. Likewise the above equation can be written as

$$\underset{A}{\lambda}(1 - CP_{AA}) = e_A + \sum_{B \neq A} \underset{B}{\lambda}\, CP_{BA}$$

Both the equations are equal, but the job transition from resource cluster A transmitted again to resource cluster A was excluded.

3.1 The arrival process into each resource cluster

The calculation of λ_A depends on the Poisson process, because the arrival process is a poisson process which helps in the distribution of various jobs across the resource clusters. For an instance, 'Z' is the random variable that represents the arrival of numbers of jobs to a resource cluster in a particular time interval. At sometimes no jobs are arriving to some of the resource clusters. In this regard, the probability of arrival of jobs as well as the probability of non arrival jobs is identified with the help of probability distribution. Two situations arise in a particular resource cluster. One is arrival of many jobs to the resource cluster and another situation is no jobs are arrived to the resource cluster in a particular time interval. The resource clusters will not be in the idle state even when no jobs are arrived. At that time jobs of the other resources are shared with the empty resource clusters to process the jobs. The time intervals of job arrivals are different as well as independent among the resource clusters which lead to the Poisson process. Poisson process is used to identify the expected value (E) of 'Z' in continuous time interval by estimating the average number of arrived jobs (λ_J) for the resource cluster. The concept of binomial distribution represents the expected value (E) of a random variable is equal to the number of trials that the random variable undergoes and the probability of success of the trials. Binomial distribution has been reformed according to the proposed resource cluster problem by specifying the number of trials as processing at independent time intervals and by specifying the probability of success as status of arrival and non arrival jobs of a resource cluster.

$$E(Z) = \lambda_J = Q_t \times P_{AJ,NJ}$$

where λ j—arrival of jobs, Q—no of trials(independent time intervals), AJ—arrived jobs and NJ—non arrived jobs.

For example, estimating number of jobs arrives per hour will leads to the number of possible trials per hour (60 min = 1 h, so 60 trials) with the probability of success per hour.

$$\Lambda_{J/h} = 60_{min/h} \times \lambda/60_{J/min}$$

If 'K' jobs arrives at a particular time interval then the probability of random variable 'Z' will be

$$P(Z = K) = \begin{bmatrix} 60 \\ K \end{bmatrix} \left[\frac{\lambda}{60}\right]^{K} \left[1 - \frac{\lambda}{60}\right]^{60-K}$$
$$(60 - k \rightarrow \text{no job arrival})$$

The trials can be done for a second also because there is a possibility of arriving more than a job in a minute. Therefore the estimation has to be done by

$$P(Z = K) = \begin{bmatrix} 3600 \\ K \end{bmatrix} \left[\frac{\lambda}{3600}\right]^{K} \left[1 - \frac{\lambda}{3600}\right]^{3600-K}$$
$$(3600 \rightarrow 3600\text{s/min})$$

The trials can be done for a millisecond also because there is a possibility of arriving more than a job in a minute. Therefore the estimation has to be done by

$$P(Z = K) = \begin{bmatrix} 3{,}600{,}000 \\ K \end{bmatrix} \left[\frac{\lambda}{3{,}600{,}000}\right]^{K} \times \left[1 - \frac{\lambda}{3{,}600{,}000}\right]^{3{,}600{,}000-K}$$
$$(3{,}600{,}000 \rightarrow 3{,}600{,}000 \text{ ms/h})$$

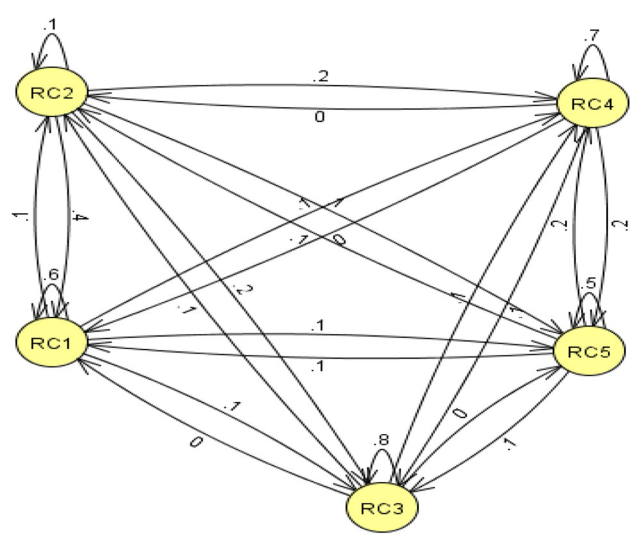
Infinite numbers of trials are performed in continuous time intervals because there is no limit for the arrival rate. It is not possible to include the limit for job arrivals in continuous markov chain. If you are doing so it will be converted in to discrete time markov process. This statement leads to reformulation of the above equation as

$$P(Z = K) = \lim_{Q\to\infty} \begin{bmatrix} Q \\ K \end{bmatrix} \left[\frac{\lambda}{Q}\right]^{K} \left[1 - \frac{\lambda}{Q}\right]^{Q-K}$$
$$\lim_{Q\to\infty} \frac{Q!}{(Q-K)!K!} \left[\frac{\lambda^K}{Q^K}\right] \left[1 - \frac{\lambda}{Q}\right]^{Q} \left[1 - \frac{\lambda}{Q}\right]^{-K}$$
$$\lim_{Q\to\infty} \left[\frac{Q(Q-1)(Q-2)\ldots(Q-K+1)}{Q^K}\right] \left[\frac{\lambda^K}{Q^K}\right]$$
$$\left[1 - \frac{\lambda}{Q}\right]^{Q} \left[1 - \frac{\lambda}{Q}\right]^{-K}$$
$$\lim_{Z\to a} B(Z)\, C(Z) = \lim_{Z\to a} B(Z) \lim_{Z\to a} C(Z)$$
$$\lim_{Q\to\infty} \frac{Q^K + \ldots}{Q^K} \frac{\lambda^K}{Q^K} \lim \left[1 - \frac{\lambda}{Q}\right]^{Q} \left[1 - \frac{\lambda}{Q}\right]^{-K}$$
$$= 1 \times \frac{\lambda^K}{K!} e^{-\lambda} \times 1$$
$$P(Z = K) = \lim_{Q\to\infty} \begin{bmatrix} Q \\ k \end{bmatrix} \left[\frac{\lambda}{Q}\right]^{K} \left[1 - \frac{\lambda}{Q}\right]^{Q_K} = \frac{\lambda^K}{K!} e^{-\lambda}$$

From the above equation the Poisson process shows the perfect arrival rate by undergoing much number of trials as a continuous process and the needed granularity has been obtained.

3.2 Distribution of jobs among the resource clusters

Markov chain is an excellent tool to identify the distribution of jobs among the resource clusters if the service times and the interarrival times of jobs are exponentially distributed. Let us consider five resource clusters RC1, RC2, RC3, RC4 and RC5 for the proposed system as they are modeled by Markov chain to identify the probability of job distribution. The continuous markov chain model creates an evaluation model to manage more number of jobs by calculating the arrival time at infinite time by using the Poisson process. (Rev 2 res 2) The continuous time markov chain process performs well at infinite time arrivals of the job in the cloud environment as compared with the discrete time markov chain process. Because in discrete time markov chain the time limit should be specified. The figure shows the transition diagram of resource cluster modeled by Markov chain that depicts the distribution of jobs among the resource clusters and identifies the probability of mean number of jobs in each and every resource cluster. Jobs will arrive internally or externally to RC1 in that 60% of jobs reside in RC1 and remaining 40% of jobs are scattered across others resource clusters such as RC2, RC3, RC4 and RC5. The jobs, residing in the resource cluster are based on the calculation of resource potential using EPRNW algorithm. 10% of jobs reside in RC2 which arrived internally or externally to RC2 and remaining 90% of the jobs are scattered across other resource clusters and same Markov property is followed for the remaining resource clusters. The transition probability matrixes show the clear representation of probability distribution of jobs and mean number of jobs in a resource cluster.



Transition probability matrix

$$P = \begin{array}{c c} & \begin{array}{ccccc} RC1 & RC2 & RC3 & RC4 & RC5 \end{array} \\ \begin{array}{c} RC1 \\ RC2 \\ RC3 \\ RC4 \\ RC5 \end{array} & \begin{pmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.4 & 0.1 & 0.2 & 0.2 & 0.1 \\ 0 & 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0 & 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.5 \end{pmatrix} \end{array}$$

Initial state distribution matrix for resource clusters. Here the State distribution matrix is used to identify the percentage or probability of device or resource utilization of the cloud environment in the job distribution.

RC1 others

$$SDC1_0 = [0.3\,0.7]$$

The initial state distribution matrix for RC1 states that RC1 holds 30% of jobs and remaining 70% of jobs are scattered across other resource clusters.

RC2 others

$$SDC2_0 = [0.2\,0.8]$$

The initial state distribution matrix for RC2 states that RC2 holds 20% of jobs and remaining 80% of jobs are scattered across other resource clusters.

RC3 others

$$SDC3_0 = [0.4\,0.6]$$

The initial state distribution matrix for RC3 states that RC3 holds 40% of jobs and remaining 60% of jobs are scattered across other resource clusters.

RC4 others

$$SDC4_0 = [0.6\,0.4]$$

The initial state distribution matrix for RC4 states that RC4 holds 60% of jobs and remaining 40% of jobs are scattered across other resource clusters.

RC5 others

$$SDC5_0 = [0.3\,0.7]$$

The initial state distribution matrix for RC5 states that 30% of jobs reside in RC5 and remaining 70% of jobs are scattered across other resource clusters. By using these state distribution matrix, probability of distribution of jobs among resource clusters at independent time interval can be calculated.

Probability of jobs distributed at RC1

$$= (0.3)(0.6) + (0.7)(0.5)$$
$$= 0.18 + 0.35 = 0.53$$

Therefore, the first state transition matrix for RC1 at independent time interval $SDC1_1 = [0.53\,0.47]$. This implies 53% of jobs from the arrived jobs will reside in RC1 and 37% of jobs will be scattered to other resource clusters.

Probability of jobs distributed at RC2

$$= (0.2)(0.4) + (0.8)(0.7)$$
$$= 0.8 + 0.56 = 0.64$$

Therefore, the first state transition matrix for RC1 at independent time interval $SDC1_2 = [0.64\,0.36]$. This implies 64% of jobs from the arrived jobs will reside in RC2 and 36% of jobs will be scattered to other resource clusters.

Probability of jobs distributed at RC3

$$= (0.4)(0) + (0.6)(0.11)$$
$$= 0 + 0.66 = 0.66$$

Therefore, the first state transition matrix for RC1 at independent time interval $SDC1_3 = [0.66\,0.34]$. This implies 66% of jobs from the arrived jobs will reside in RC2 and 34% of jobs will be scattered to other resource clusters.

4 Results and discussions

The experimental graphs below depict the performance evaluations of the existing estimating model and the proposed Markov chain along with the Poisson model. In the proposed system, various number of resource clusters (RC = 50, 100) are used to perform the task in the cloud environment. Figures 1 and 2 shows the calculation of arrival rate of jobs at 50 and 100 resource clusters by Existing estimated Model (EEM) and Markov chain Poisson model (MCP). The MCP model identifies and schedules more number of arrivals of jobs than the EEM.

The percentage probability of arrival rate for exponential values has been calculated and the comparison table proves that the MCP identifies more number of arrivals even in exponential time and the percentage probability is increased in MCP regarding to the identification of arrival rate of the jobs.

EEM				MCP			
Percentage of arrival rates	Expected exponential percentage	Lower percentage	Upper percentage	Percentage of arrival rates	Expected exponential percentage	Lower percentage	Upper percentage
84.36	14.28571	0.38242	1.89471	84.4994	13.08771	1.58563	7.85604
84.4319	28.57143	0.76872	3.80865	84.8917	28.56718	3.17125	15.71209
84.7533	42.85714	1.15898	5.74222	84.9627	47.51252	4.75688	23.56813
84.8346	57.14286	1.55329	7.69583	84.9727	71.93731	6.3425	31.42418
84.8627	71.42857	1.95173	9.6699	84.9857	106.36212	7.92813	39.28022
84.93	85.71429	2.35438	11.66486	85.1	165.21173	9.51376	47.13626

Figures 3 and 4 shows the calculation of percentage probability of arrival rate of jobs at 50 and 100 resource clusters by Existing Estimated Model (EEM) and Markov chain Poisson model (MCP).

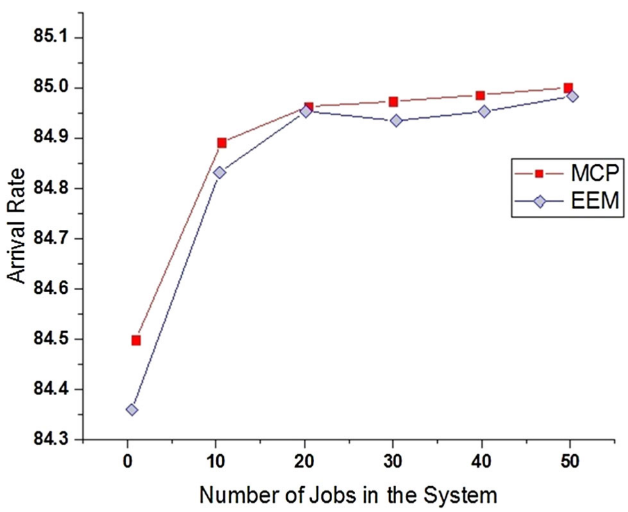


Fig. 1 RC = 50

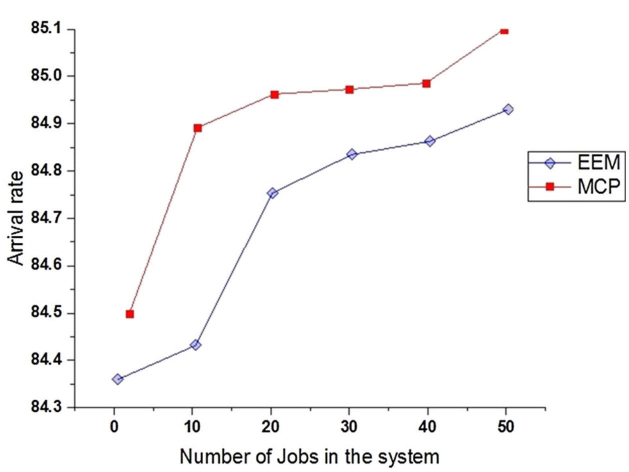


Fig. 2 RC = 100

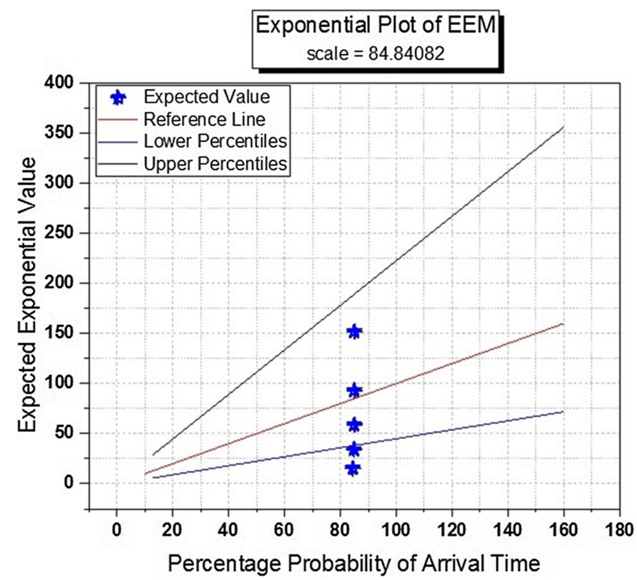


Fig. 3 % Probability of arrival rate at RC = 50

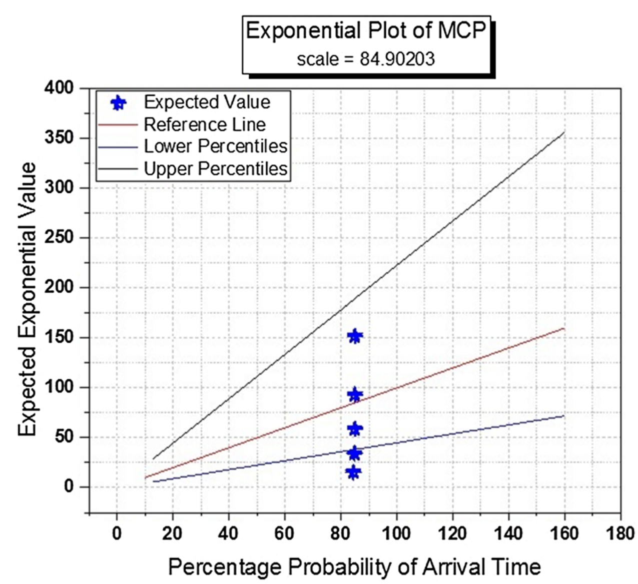


Fig. 4 Probability of arrival rate at RC = 100

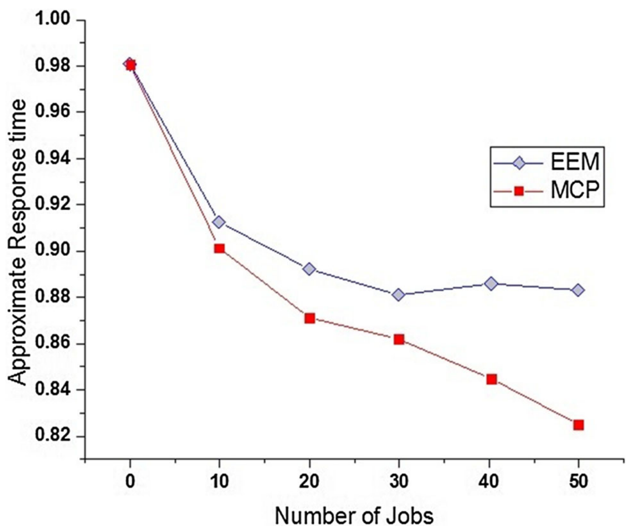


Fig. 5 Calculation of response time at 50 RC

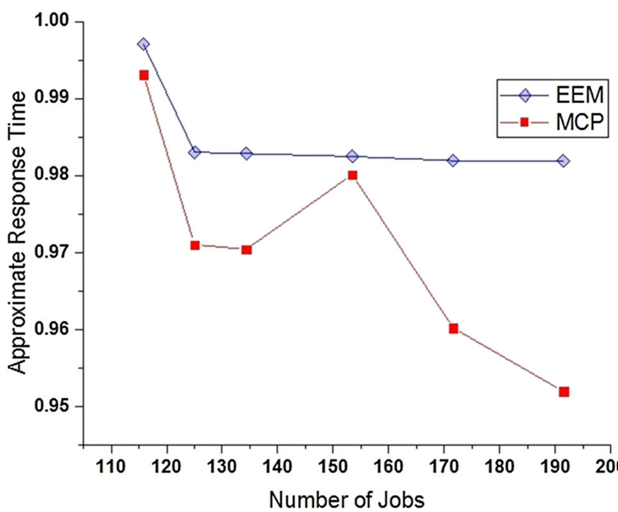


Fig. 6 Calculation of response time at 100 RC

Figures 5 and 6 shows the calculation of response time of jobs at 50 and 100 resource clusters by existing estimated model (EEM) and Markov chain Poisson model (MCP). The MCP model minimizes the response time of jobs and identifies the mean number of jobs present in the cloud environment and ensures better resource utilization. By minimizing the response time, the makespan of the cloud environment is reduced while scheduling the jobs across the resource clusters in the cloud environment.

Figures 7 and 8 shows the calculation of percentage probability of response time of jobs at 50 and 100 resource clusters by existing estimated model (EEM) and Markov chain Poisson model (MCP).

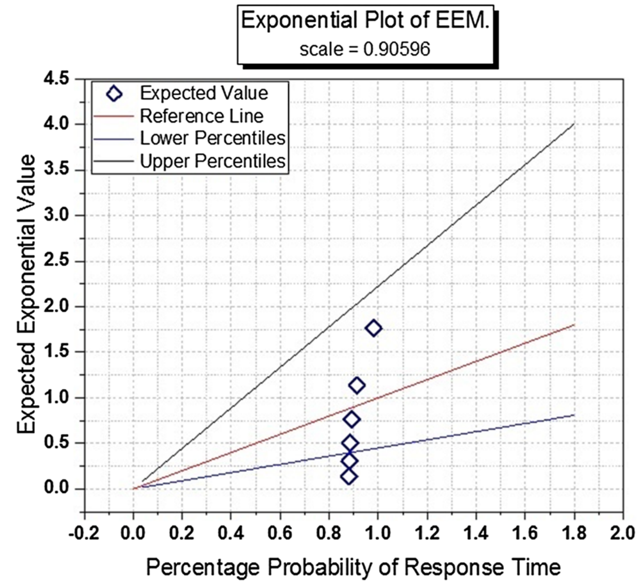


Fig. 7 % Probability of response time at RC = 50

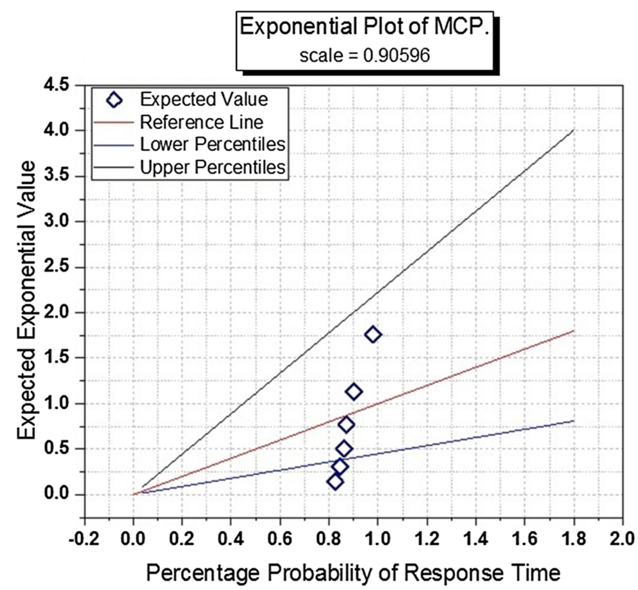


Fig. 8 % Probability of response time at RC = 50

The percentage probability of response time for exponential values has been calculated and the comparison table proves that the MCP minimizes the response time of jobs even in exponential time and the percentage probability of responding jobs are increased in MCP when comparing to EEM.

EEM				MCP			
Response time	Expected exponential percentage	Lower percentage	Upper percentage	Percentage of arrival rates	Expected exponential percentage	Lower percentage	Upper percentage
0.88106	10.9375	0.00409	0.02027	0.95199	10.9886	0.91256	0.96234
0.88319	26.5625	0.00822	0.04074	0.96024	26.8765	0.91876	0.96404
0.88593	42.1875	0.0124	0.06142	0.97044	42.6328	0.92266	0.96515
0.8922	57.8125	0.01662	0.08232	0.97103	58.4231	0.92556	0.96602
0.91251	73.4375	0.02088	0.10344	0.98013	74.7653	0.92791	0.96674
0.9809	89.0625	0.02518	0.12478	0.99312	90.1852	0.92989	0.96737

A multiobjective graph is represented in Fig. 9. It shows the probability of response time of the resource clusters, cumulative percentage of jobs in resource clusters that ensures resource utilization of the cloud environment and exponentially distributed arrival rate. As the response time, service time and completion time minimized, the makespan of the cloud environment also minimized. This is due to the definition of the makespan. Makespan is the time taken by an algorithm or a procedure to complete all its jobs of the cloud environment. The difference between the service time and completion time suggest that service time is the inner hole of the completion time. Service time is the time taken to complete the particular job of the environment. The completion time includes both the response time as well as the service time.

5 Conclusion

As for as the donors and consumers of the cloud concern, the performance evaluation of resources plays a vital role. A mathematical model on continuous time Markov chain regarding to the identification of resource utilization, service time are combined with the Poisson process regarding to the arrival rate of the jobs are proposed to produce an efficient performance evaluation model for the cloud environment resource clusters. The proposed mathematical model performs well in identifying exponential arrival rate, response time and resource utilization for different types of resource clusters in a cloud environment. Mathematical experiments have been carried out to evaluate the proposed system. An efficient performance evaluation model which produces near optimal solution for a cloud environment have been attained by identifying the arrival rate of jobs in exponential time, probability of response time of the resource clusters and cumulative percentage of jobs in the resource clusters. Makespan of the particular cloud environment have been minimized with the help of the proposed model which results in an accurate identification of arrival rate as well as the resource utilization. In future, proposed model can be extended by comparing discrete time Markov chain with continuous time markov chain for evaluating the probability of the performance of the resource clusters and conversion of discrete time markov chain to continuous time markov chain and vice versa will be the further extent. The conversion is for identifying the job arrivals at discrete time intervals and provide limit for the trials which will be adequate even for small environments.

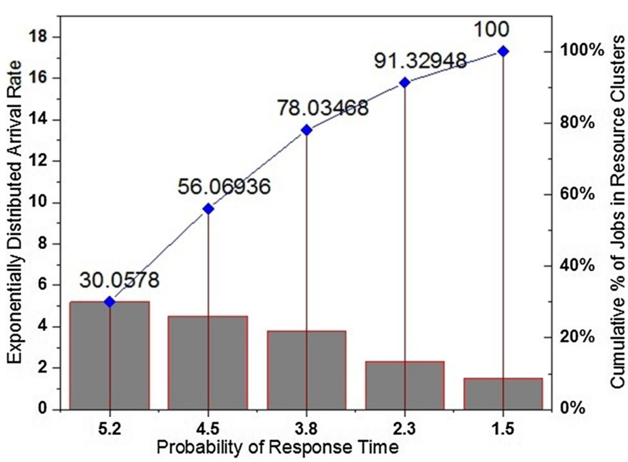


Fig. 9 Identification of arrival rate, response time, Resource Utilization

References

1. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. J. Internet Serv. Appl. **10**(1), 7–18. ISSN 1869-0238 (2010)
2. Kowsigan, M., Balasubramanie, P.: Scheduling of jobs in cloud environment using soft computing techniques. Int. J. Appl. Eng. Res. **10**(38). ISSN 0973-4562 (2015)
3. Kowsigan, M., Balasubramanie, P.: An improved job scheduling in cloud environment using auto associative memory network. Asian J. Res. Soc. Sci. Humanit. **6**(12), 390–410. ISSN 2249-7315 (2016)
4. Karlin, S., Taylor, H.M.: A First Course in Stochastic Processes, 2nd edn, pp. 221–228. Academic Press, Cambridge (1975)
5. Harchol-Balter, M.: Performance Modeling and Design of Computer Systems-Queuing Theory in Action, pp. 106–118. Cambridge University Press, Cambridge (2013)

6. Khazaei, H., Misic, J., Misic, V.B.: Performance analysis of cloud computing centers using M/G/m/m+r queuing systems. IEEE Trans. Parallel Distrib. Syst. **23**(5), 936–943 (2012)
7. Jain, S., Jain, S.: Analysis of multi level feedback queue scheduling using markov chain model with data model approach. Int. J. Adv. Netw. Appl. **07**(06), 2915–2924. ISSN 0975-0290 (2016)
8. Rabbani, M., Tanhaie, F.: A Markov chain analysis of the effectiveness of drum-buffer-rope material flow management in job shop environment. Int. J. Ind. Eng. Comput.**6**(4), 457–468. ISSN 1923-2934 (2015)
9. Terefe, M.B., Lee, H., Heo, N., Fox, Geoffrey C., Sangyoon, Oh: Energy-efficient multisite offloading policy using Markov decision process for mobile cloud computing. Pervasive Mob. Comput. **27**, 75–89 (2016)
10. Barnes, A.K., Balda, J.C., Escobar-Mejía, A.: A Semi-Markov model for control of energy storage in utility grids and microgrids with PV generation. IEEE Trans. Sustain. Energy **6**(2), 546–556. ISSN 1949-3037 (2015)
11. Zhang, W., Wen, Y., Wu, D.O.: Energy-efficient scheduling policy for collaborative execution in mobile cloud computing. In: Proceedings IEEE INFOCOM. ISSN: 0743-166X (2013)
12. Park, J.S., Yu, H.C., Chung, K.S., Lee, E.Y.: Markov chain based monitoring service for fault tolerance in mobile cloud computing. In: IEEE Workshops of International Conference on Advanced Information Networking and Applications (2011)
13. Wang, J., Qiu, M., Guo, B.: High reliable real-time bandwidth scheduling for virtual machines with hidden Markov predicting in telehealth platform. Future Gener. Comput. Syst. **49**, 68–76 (2015)
14. Yin, J., Lu, X., Zhao, X., Chen, H., Liu, X.: BURSE: a bursty and self-similar workload generator for cloud computing. IEEE Trans. Parallel Distrib. Syst. **26**(3). ISSN 1045-9219 (2015)
15. Wu, H., Knottenbelt, W., Wolter, K.: Analysis of the energy-response time tradeoff for mobile cloud offloading using combined metrics. In: IEEE 27th International Teletraffic Congress (2015)
16. Zhang, R., Su, X., Wang, J., Wang, C., Liu, W., Lau, R.W.H.: On mitigating the risk of cross-VM covert channels in a public cloud. IEEE Trans. Parallel Distrib. Syst. **26**(8), 1. ISSN 1045-9219 (2015)
17. Peng, Z., Cu, D., Zuo, J., Li, Q., Bo, Xu, Lin, Weiwei: Random task scheduling scheme based on reinforcement learning in cloud computing. Cluster Comput. **18**(4), 1595–1607 (2015)
18. Lu, C.-T., Chang, C.-W., Li, J.-S.: VM scaling based on Hurst exponent and Markov transition with empirical cloud data. J. Syst. Softw. **9**, 199–207 (2015)
19. Hyytiä, E., Spyropoulos, T., Ott, J.: Offload (only) the right jobs: robust offloading using the Markov decision processes. In: IEEE 16th International Symposium on a World of Wireless, Mobile and Multimedia Networks (2015)
20. Zhang, Q., Chen, Z., Yang, L.T.: A nodes scheduling model based on Markov chain prediction for big streaming data analysis. Int. J. Commun. Syst. **28**(9), 1610–1619. ISSN 1099-1131 (2014)
21. Kowsigan, M., Rajkumar, S., Seenivasan, P., Vikram Kumar, C.: An enhanced job scheduling in cloud environment using improved metaheuristic approach. Int. J. Eng. Res. Technol. **6**(2), 184–188 (2017)
22. Kowsigan, M., Kalicharan, S., Karthik, P., Manikandan, A., Manikandan, R.: An enhanced job scheduling in cloud environment using probability distribution. Int. J. Eng. Technol. **9**(2), 1374–1381 (2017)

M. Kowsigan received his B.Tech. degree in Information Technology from Anna University in 2008 and M.Tech. Information Technology in Anna University in 2011. He is pursuing his Ph.D. in Anna University. He is currently an Assistant Professor in Department of Information Technology in Sri Krishna College of Technology, Coimbatore. He have published many papers in many reputed journals and conferences. His main research interest includes Cloud Computing, Grid Computing, Soft Computing and Internet of Things.

P. Balasubramanie currently working as a Professor in the Department of Computer Science & Engineering, Kongu Engineering College, Tamilnadu, India. He is one of the approved supervisor of Anna University Chennai and guided 26 Ph.D. scholars. Currently he is guiding 10 Ph.D. scholars. He has published 203 articles in national/international journals. He has published 11 books with the reputed publishers. Three of the books published are used as text/reference books by many of the leading universities in India. He has completed one AICTE research promotion scheme (RPS) as a Principal Investigator. Currently he is working in a UGC minor research project as a principal investigator. He has received 13 lakhs grant from various funding agencies like AICTE, CSIR, DRDO, NBHM, ICMR and so on and organised 21 SDP/STTP/Seminar/workshops for the benefit of faculty members and research scholars. So far he has received 14 awards. He is actively involved in training and consultancy activities. His field of specialization includes Cloud Computing, Datamining and Networking.