Formulation of Profit Aware Algorithms for VM Provisioning using Finite Queueing Model

N.Neelima, B.Basaveswar Rao, K.Gangadhara Rao K.Chandan

Abstract: ***This paper analyses the profit model and suggests two auto scaling algorithms for web applications running on cloud using analytical multi sever finite Queueing model. The profit function is formulated and numerically illustrated with two relevant scenarios that will have both qualitative and quantitative bearing on VM provisioning. The important QoS metrics like blocking probability and waiting time in the queue are taken in to consideration. Based on these metrics the profit trend analysis is done and then subsequently the algorithms are used to strategize VM Provisioning. Finally the execution flow of the algorithms and conclusions are presented.***

Index terms: Cloud Computing, Queueing theory, Profit analysis, VM provisioning.

I. INTRODUCTION

Nowadays, Cloud Computing has become an essential portion of the digital world for organisations and enterprises. However the cloud service provider's challenge is to make maximum profit by serving cloud customers with in stipulated time. Cloud consists of a group of hardware, storage, interface and network which enable the computing as a service to be delivered. Cloud customers instead of purchasing computing infrastructure and installing software, they can access their data worldwide from any computer as long as the Internet connection is available. Because of the reduced cost, lower investment in IT infrastructure, and sufficient resource availability, cloud computing is becoming ever more attractive option to run applications. The web application hosting is playing a significant role to promote the business applications and services. The web application hosting mainly dependent on elastically allotted resources and application deployment. For profit maximization, a service provider should understand both hosting charges and business costs, and how they are determined by the features of the applications and the configuration of a multiserver system

Cloud computing provides various services to their customers based on the Service Level Agreement (SLA). SLA is an agreement used to guarantee web service delivery governed by the terms and conditions agreed up on by both the service provider and consumer [4].The important part between the provider and customer is the rapport to be preserved in terms of price for which they must agree. Pricing represents an important indicator for the success of companies which provide services or products [3].The communication and negotiation about the price between provider and customer is based on various Quality of Service (QoS) metrics. Each provider has his structure (accounting system) for calculating the cloud services offered for customers. The Cloud Service Provider (CSP) objective is to maximize profit by orchestrating the services in professional way, while each customer's goal is to have the best service at a competitive price. The economics of cloud computing is an essential issue .Thus, satisfying both parties requires an optimal pricing methodology. The price charged is one of the most important metrics that a service provider can control to encourage the usage of its services [5].

The approach of pricing can be categorized into two models such as fixed pricing model and dynamic pricing model. In fixed pricing model the prices of various resources are determined prior by the provider and price charging cannot change for the contractual period. Fixed pricing contains the pay per use pricing and subscription pricing [6]. Fixed pricing is more straightforward and simple to use, but it would not be fair to all users, because it is not necessary, that all users have similar needs [7]. Fixed pricing also restricts the providers profit whenever the demand either increases or decreases. In Dynamic pricing model prices change dynamically according to market dynamics based on the variations in demand and supply [8] [9].

For profit maximization, the CSP must focus on the factors like the amount of service, the workload of an application environment, the configuration of a multiserver system, the service-level agreement, satisfaction of consumer, server speed, and power consumption etc.

To increase the revenue the CSP will construct and pin together a multiserver system with several servers of high speed. Since the particular service time (i.e., the task response time) contains task waiting time and task execution time, a lot of servers shrink the waiting time and quicker servers scale back each waiting time and execution time. However, a lot of servers (i.e., a bigger multiserver system) increase the price of facility dealings from the infrastructure vendors and also the cost of base power consumption. Furthermore, quicker servers increase the price of energy consumption. Such increased price might counterweight the gain from penalty reduction.

Revised Manuscript Received on December 22, 2018.

N.Neelima, Assistant Professor, Velagapudi Ramakrishna Siddhartha Engineering College,A.P.
B.Basaveswar Rao, Acharya Nagarjuna University Guntur, A.P
K.Gangadhara Rao, Acharya Nagarjuna University Guntur, A.P
K.Chandan, Acharya Nagarjuna University Guntur, A.P

Some of the researchers have done research on various pricing schemes using M/M/m queueing model [23][28].Some researchers have worked on how the profit model affect in different resources, their comparison, also the pricing model for two important cloud platforms namely, 1) Google Cloud Computing; and 2) Amazon Web Service [29]. Thus far no researcher has considered the relationship between the profit trends and the SLA. This paper tries to explore the relationship between the profit trends and SLA in the context of maximization of the profits for the CSP without compromising on the parameters of SLA. To achieve the aforementioned objective this paper proposes an analytical model to analyse the profit trends for the cloud provider using M/M/S/K queuing model [30] with static pricing. The profit model based on the performance metrics like blocking probability and average waiting time in the queue are taken in to consideration because these metrics are influential factors for entering in to SLA. The profit model is formulated and analysed with two scenarios. The main objective of this paper is as follows:

(i)To formulate a profit based queueing model for Web Application running on Cloud
(ii)The model is analysed for different values of user requests, number of virtual Servers and for different buffer size.
(iii)Based on these analysis two algorithms for VM provisioning/releasing have been formulated.
(iv)Finally the flow of these algorithms is discussed.

The rest of the paper is organised as follows, section 2 discusses about related work, and in section 3 Proposed Profit Model is presented, the Numerical Illustration is carried out in section 4 , VM provisioning algorithms based on Profit trend analysis are given in section 5 . Finally conclusions are drawn in section 6.

II. RELATED WORK

In order to make cloud computing pricing facility more attractive many of the researchers have made several proposals. Amazon provides on-demand instance services which let you pay for compute capacity by the hour with no long-term commitments. This sets you free from the prices and complexities of scheduling, buying, and maintaining hardware and transforms what are commonly large fixed costs into much smaller variable costs [10]. In [11] Dynamic resource pricing on federated clouds was proposed by which the price can be determined based on the level of supply and demand. . In [12], authors employed Competition-based pricing model, by which provider sets the price according to competitors. Pricing algorithm for cloud computing resources that could be used for minimizing cost as well as maximizing profits for the service provider was introduced in [13]. Customer-based pricing model in [14] was introduced that; the price could be specified according to the customers' needs (what the consumer is ready to pay). Macias and Guitart [15] introduced a simulation based genetic model for pricing in cloud computing markets. The results of the simulation illustrated that service providers using genetic pricing attained revenues up to 100% greater than the other dynamic pricing approaches and up to 1000% more than the fixed pricing strategy.

A dynamic model in which cloud provider uses job scheduling mechanism to set resources and prices was proposed in [19].A dynamic approach in which the cloud provider specifies the profit level to set of resources and services depending up on the real time market by mutually maximizing revenues and minimizing electricity costs was introduced in [20,21]. Pal and Hui [22] have considered an economic model for fixing prices of resources. They used game theory and presented several economic models. An optimization problem is formulated and solved analytically by using M/M/c queuing model considering a multiserver system in [23]. Their pricing model took into considerations the amount of a service, the workload of an application environment, the configuration of a multiserver system, the service-level agreement, and the satisfaction of a consumer. In order to guarantee the quality of service requests and maximize the profit of service providers, the authors proposed a novel double-quality guaranteed renting scheme for service providers considering as an $M/M/c+D$ queuing model in[24].In [25] the authors developed a profit function, in which both the system blocking loss and the user abandonment loss are evaluated in total revenue by considering cloud data center as a queuing system with finite capacity, interarrival and service times were both assumed to be exponentially distributed.

The auto-scaling process of web applications which dynamically adapt the resources assigned to the web applications, depending on the input workload was proposed in [16, 17]. Cost-Optimized Resource Provision for Cloud Applications for High Performance Computing and Communications was proposed in [18].

III. PROFIT MODEL

In general, various factors influence the profit of the cloud provider. These include power consumption cost, service charge, infrastructure cost and VM's operating cost etc. In this paper M/M/S/K Queueing model with cost analysis is adopted based on [26][27][31]. The cost model is explained and analytically derived in the previous work [26] and validated with experimental study through AWS in [27].

Notations:

λ= the arrival rate of user requests,
μ= the service rate of each virtual machine for completion of user request
S= Number of Virtual machines
K=Buffer capacity
P_k = Blocking Probability (system is full)
L_q= Average Customers in Queue
λ_e =Effective Arrival Rate
W_q=Average waiting time in the Queue
γ=the unit of charge for each user request
α= The unit of cost for request completion to consume the resources

β= The unit of cost for request waiting in the buffer to consume the resources

The profit (P) function is defined as follows:

$$P = \gamma\lambda_e - \alpha s\mu - \beta L_q$$
$$= \gamma\lambda_e - \alpha\, s\, \mu - \beta\, \lambda_e W_q$$
$$= \gamma\lambda\,(1-P_k) - \alpha\, s\, \mu - \beta\,(1-P_K)\,\lambda\, W_q \quad (1)$$

Where $L_q = \lambda_e W_q$ and $\lambda_e = (1-P_K)\,\lambda$

The profit model consists of three parts, they are unit of money realised for each request is the Revenue (R), cost incurred for the requests actual service completion by the VM's is Operational Cost (OC) and the number of requests waiting in the queue is the cost of waiting cost (WC).The R depends on λ .The OC is independent of λ and is dependent on S and μ which are elastic by virtue of the inherent characteristics of elasticity. The WC depends upon all the parameters λ, S, μ and K. Thus the profit function P=R-OC-WC.

IV. NUMERICAL ILLUSTRATION

The objective of this section is to analyse the impact of the profit based on the different values of λ, S, K and fixed values of α, β, γ and μ as 100 ms as the Web server was structured to yield a webpage of size 580 bytes when receiving a request .The server was adjusted until we achieve the CPU utilization of up to 100% when it receives 100 Requests/sec to match the capacity of the target medium sized virtual machine [32] [33]. The above profit equation (1) has been justified by the numerical illustration with two scenarios (i) S and λ varying scenario (ii) K and λ varying scenario. For both the scenarios α, β, γ are taken as 10, 5 and 35 units respectively and γ must be greater than α+ β. The value of the profit rounded to near 10.

A.S and λ varying scenario:

In this scenario to identify the behaviour of the profit when λ, S increases are shown in Table 1 for K=30

Table 1: Effect of Profit in S and λ varying scenario

S / λ	2	4	6	8	10
200	4703	2999	999	-1000	-3000
400	4865	9460	7997	5999.7	4000
600	4862	9879	14190	12991	10999
800	4861	9875	14892	18885	17966
1000	4861	9873	14887	19888	23538
1200	4861	9872	14885	19900	24841
1400	4860	9872	14884	19897	24908
1600	4860	9871	14883	19895	24908
1800	4860	9871	14882	19894	24906
2000	4860	9871	14882	19893	24905
2200	4860	9871	14882	19893	24904
2400	4860	9871	14882	19892	24904
2600	4860	9870	14881	19892	24903
2800	4860	9870	14881	19892	24903
3000	4860	9870	14881	19892	24902
3200	4860	9870	14881	19892	24902
3400	4860	9870	14881	19892	24902
3600	4860	9870	14881	19891	24902
3800	4860	9870	14881	19891	24902
4000	4860	9870	14881	19891	24902

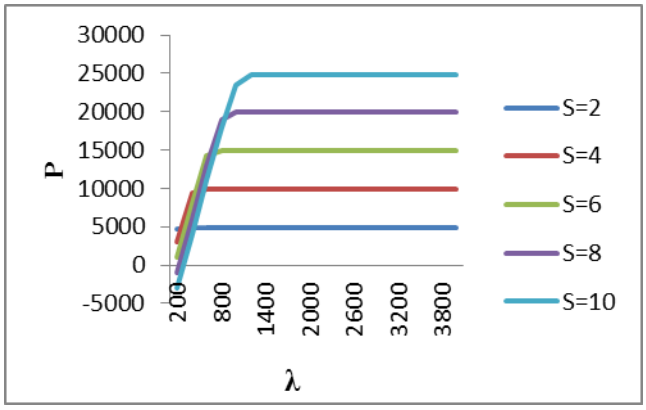

Figure 1: Effect of Profit in S and λ varying scenario

From the Table 1 as well as above Figure 1, it is observed that as the λ increases the profit also increases up to certain value of λ then profit stabilizes .This phenomena follows for different values of virtual servers. The main observation of this scenario is that when λ and S increases the profit stabilization point also increases.

B.K and λ varying scenario:

In this scenario to observe the changes in profit by varying K and λ are shown in Table 2 for fixed value of S=10.

Table 2: Effect of Profit in K and λ varying scenario

K / λ	10	20	30
200	-3000.3	-3000	-3000
400	3925.7	4000	4000
600	10094	10994	10999
800	14593	17739	17966
1000	17490	22594	23538
1200	19319	24423	24841
1400	20513	24843	24908
1600	21329	24930	24908
1800	21911	24948	24906
2000	22343	24953	24905
2200	22673	24953	24904
2400	22934	24953	24904
2600	23144	24953	24903
2800	23316	24953	24903
3000	23460	24952	24902
3200	23581	24952	24902
3400	23685	24952	24902
3600	23776	24952	24902
3800	23854	24952	24902
4000	23924	24952	24902

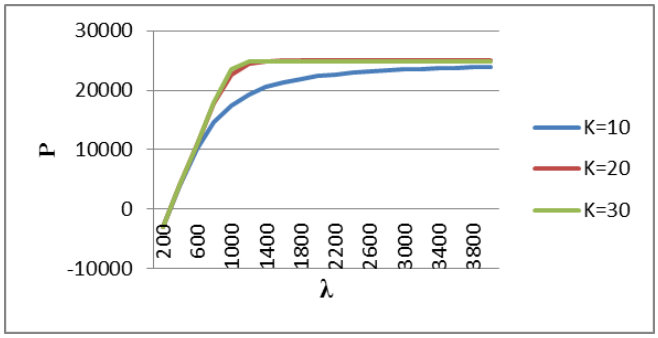

Figure 2: Effect of Profit in K and λ varying scenario

By observing Table 2 and Figure 2 it is noted that the profit increases when λ increases and then it stabilizes for different values of K. For K>=20 there is no significant difference in profit i.e. for the values of K>=20 the buffer impact on profit is negligible.

It is concluded that by examining the two scenarios numerical results, the S and λ varying scenario has more effect on profit whereas K and λ varying scenario is comparatively ineffective, and for all the values of K>=20 the difference in profit is insignificant .By this observation ,it is motivated to observe the profit trend in depth by adapting S and λ varying scenario for more values of λ(from 100 to 4000) and S(2 to 30) to design algorithms for VM provisioning based on profit trend analysis. These results are presented in Table 3 as well as in Figure 3.In the next section; the VM provisioning algorithms are presented. Though both the scenarios are important the first one contributes to guide the making of SLA.

Table 3: Effect of Profit trend for more values of λ and S in S and λ varying scenario

	S=2	S=4	S=6	S=8	S=10	S=12	S=14	S=16	S=18	S=20	S=22	S=24	S=26	S=28	S=30
λ=100	1498.3	-500	-2500	-4500	-6500	-8500	-10500	-12500	-14500	-16500	-18500	-20500	-22500	-24500	-26500
200	4703.9	2999.1	999.96	-1000	-3000	-5000	-7000	-9000	-11000	-13000	-15000	-17000	-19000	-21000	-23000
300	4870	6491.6	4499.5	2500	499.998	-1500	-3500	-5500	-7500	-9500	-11500	-13500	-15500	-17500	-19500
400	4865	9460.8	7997.1	5999.7	4000	2000	-0.0002	-2000	-4000	-6000	-8000	-10000	-12000	-14000	-16000
500	4863.3	9884.5	11464	9498.6	7499.8	5500	3500	1500	-500	-2500	-4500	-6500	-8500	-10500	-12500
600	4862.5	9879.9	14190	12991	10999	9000	7000	5000	3000	1000	-1000	-3000	-5000	-7000	-9000
700	4862	9876.7	14858	16375	14496	12500	10500	8500	6500	4500	2500	500	-1500	-3500	-5500
800	4861.7	9875	14892	18885	17966	15998	14000	12000	10000	8000	6000	4000	2000	-0.00004	-2000
900	4861.4	9874	14890	19754	21203	19484	17498	15500	13500	11500	9500	7500	5500	3500	1500
1000	4861.2	9873.3	14887	19888	23538	22890	20990	18998	17000	15000	13000	11000	9000	7000	5000
1100	4861.1	9872.9	14886	19901	24562	25943	24442	22492	20498	18499	16500	14500	12500	10500	8500
1200	4861	9872.5	14885	19900	24841	28140	27735	25959	23991	21997	19999	17999	16000	14000	12000
1300	4860.9	9872.2	14884	19898	24899	29281	30591	29336	27463	25489	23496	21498	19499	17499	15499
1400	4860.8	9872	14884	19897	24908	29720	32677	32473	30872	28960	26984	24992	22995	20996	18996
1500	4860.8	9871.8	14883	19896	24909	29862	33905	35139	34127	32379	30450	28474	26483	24487	22488
1600	4860.7	9871.7	14883	19895	24908	29903	34503	37131	37081	35682	33865	31929	29954	27965	25968
1700	4860.7	9871.5	14883	19894	24907	29915	34760	38424	39568	38771	37181	35330	33390	31415	29424
1800	4860.6	9871.4	14882	19894	24906	29918	34863	39166	41480	41528	40326	38635	36763	34817	32835
1900	4860.6	9871.3	14882	19894	24906	29918	34903	39557	42819	43850	43216	41788	40033	38138	36174
2000	4860.6	9871.3	14882	19893	24905	29917	34918	39752	43685	45687	45772	44727	43156	41344	39408
2100	4860.5	9871.2	14882	19893	24905	29917	34924	39847	44214	47056	47939	47393	46082	44393	42501
2200	4860.5	9871.1	14882	19893	24904	29916	34926	39893	44525	48025	49701	49744	48772	47251	45419
2300	4860.5	9871.1	14882	19893	24904	29915	34927	39915	44704	48687	51083	51759	51193	49886	48134
2400	4860.5	9871	14882	19892	24904	29915	34926	39926	44806	49126	52133	53443	53333	52282	50630
2500	4860.4	9871	14882	19892	24903	29915	34926	39932	44865	49414	52913	54819	55190	54432	52897
2600	4860.4	9870.9	14881	19892	24903	29914	34926	39934	44898	49600	53481	55924	56780	56339	54938
2700	4860.4	9870.9	14881	19892	24903	29914	34925	39935	44918	49721	53892	56799	58125	58016	56762
2800	4860.4	9870.8	14881	19892	24903	29914	34925	39935	44929	49800	54187	57485	59252	59480	58384
2900	4860.4	9870.8	14881	19892	24903	29914	34925	39935	44936	44936	54398	58021	60191	60753	59821
3000	4860.4	9870.8	14881	19892	24902	29913	34924	39935	44940	49885	54550	58437	60970	61856	61092
3100	4860.3	9870.7	14881	19892	24902	29913	34924	39935	44942	49907	54659	58760	61614	62809	62215
3200	4860.3	9870.7	14881	19892	24902	29913	34924	39935	44943	49922	54738	59011	62147	63634	63209
3300	4860.3	9870.7	14881	19892	24902	29913	34924	39935	44944	49932	54795	59207	62588	64348	64089
3400	4860.3	9870.7	14881	19892	24902	29913	34923	39934	44944	49939	54837	59360	62954	64966	64871
3500	4860.3	9870.6	14881	19891	24902	29913	34923	39934	44944	49944	54868	59480	63257	65503	65566
3600	4860.3	9870.6	14881	19891	24902	29912	34923	39934	44944	49947	54891	59575	63510	65970	66187
3700	4860.3	9870.6	14881	19891	24902	29912	34923	39934	44944	49949	54908	59649	63721	66377	66743
3800	4860.3	9870.6	14881	19891	24902	29912	34923	39934	44944	49951	54921	59709	63897	66734	67242
3900	4860.3	9870.6	14881	19891	24902	29912	34923	39933	44944	49952	54931	59756	64046	67046	67691
4000	4860.3	9870.6	14881	19891	24902	29912	34923	39933	44944	49952	54938	59794	64171	67321	68097

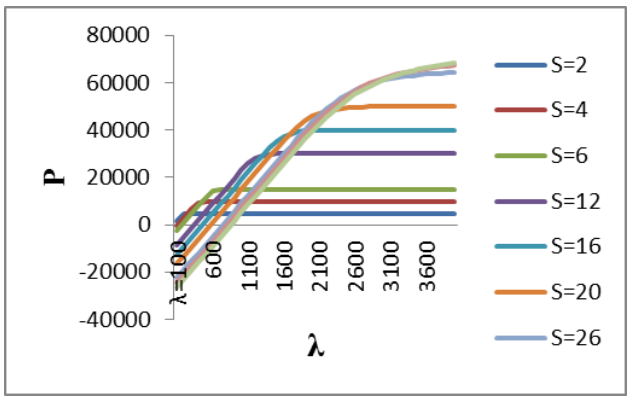

Figure 3: Effect of Profit trend for different value of S and λ

By observing the above Table 3 and Figure 3, the trend of the profit increases and then stabilizes for all values of S(2 to 30) and fixed values of μ=100 ms and K=30 . This trend behaviour can be classified in to three cases. They are i) Loss profit trend for all values of $\lambda < S\mu$ ii) profit increasing trend up to $\lambda = S\mu$ iii) profit stabilizing trend for all values of $\lambda > S\mu$. From these three types of cases, two novelistic VM provision algorithms are proposed based on profit trend

V. VM PROVISIONING ALGORITHMS BASED ON PROFIT TREND ANALYSIS

analysis. The first algorithm does not consider any QoS metrics. The second algorithm considers the two QoS Metrics i.e. blocking probability and the waiting time of user requests for provisioning VM's.

The following notations are defined for VM Provisioning algorithms:

RT=Running Time
TT=Total Time
ST=System Time
λ_c=Current Arrivals
λ_p=Previous Arrivals
P_c=Current profit
P_p=Previous profit
S_c=Current Running VM's
S_r=Required VM's
RVM =Required Virtual Machines

Algorithm 1: **Adaptive VM provisioning without QoS metrics**

Data: α
Data: β
Data: γ
Data: μ
Data: TT
Data: MinS=2
Data: MaxS=30
Result: RVM (Required Virtual Machines)
Initialization:

```
1  TT=ST
2  λc=0
3  Read Sc
4  While (True)
5        Read λc
6        TT=ST
7        Repeat
8       If λc >Scμ
9          Sr= ((λc-Scμ)/μ) +Sc
10            If Sr>MaxS
11           Sr=MaxS
12        Endif
13         Endif
14         If λc<Scμ
15            Sr = Sc-((Scμ-λc)/μ)
16        If Sr<MinS
17           Sr=MinS
18       Endif
19        Endif
20        Until ST<= TT+5 minutes
21               RVM=Sr
22           VM provisioning as per RVM
23           Sc=RVM
24   Endwhile
```

In Algorithm 1 without using Queueing Metrics, if the current user's arrival rate λ_c is more and the current running VM's are not sufficient to meet the demand then the number of VM's are scaled up as per the pseudo code mentioned above 9 to 11. Statements 12 to 14 specifies the number of current running VM's to be scale down when there is over provisioning of VM's . The algorithm repeats the process of provisioning of VM's for every 5 minutes based on system time.

Algorithm 2: Adaptive VM provisioning with QoS metrics and sustainable profit

Data: α
Data: β
Data: γ
Data: μ
Data: K
Data: RT
Data: TT
Data: MinS=2
Data: MaxS=30
Result: RVM (Required number of Virtual Machines able to meet QoS and sustainable profit)
Initialization:

```
1  TT=ST
2  Pc=0, Pp=0,  λc=0
   Read Sc
3     While (True)
4  Read λc
5  Calculated Pk, Wq
6  Pc=γλc (1-Pk) - α Sc μ - β (1-PK) λc Wq
7  TT=ST
8          Repeat
9          If (Pc<0) and (λc <Sc μ)
10                Sr= Sc-((Scμ-λc)/μ)
11                If Sr < MinS
12             Sr = MinS
13                endif
14             endif
15                 If (Pc>0) and (λc>Scμ)
16                Sr=λc/μ
17                If Sr > MaxS
18                       Sr = MaxS
19                endif
20             endif
21                 If (Pc==Pp) and (λc>Scμ)
22                Sr=λc/μ
23                      If Sr > MaxS
24             Sr =MaxS
25                endif
26             endif
27             Until ST<= TT+5 minutes
28             RVM=Sr
29             VM provisioning as per Sr

30              Pp=Pc
   Sc=RVM
```

In the Algorithm 2, statements 9 to14 represents the occurring of case (i) with loss profit because of over provisioning of VM's. So the VM's are scaled down as per required VM's. The system enters in case (ii) through statements 15 to 20 ,when profit is in increasing trend but sufficient number of VM's are not provided for achieving maximum profit . So the required VM's are scaled up with constraint of configured maximum VM's. And the system moves to case (iii), through the statements
16 to20 when the profit will not change when λ increases. If the VM's are scaled up to meet the additional values of λ requests for getting additional profit with a constraint of configured maximum VM's.

The above algorithm repeats every 5 minutes for taking VM's provisioning decision with observation of λ_c, S_c and then calculated queueing metrics P_k and W_q as QoS satisfied metrics after then calculates profit. However the algorithm time limit can be changed as per choice.

The analytical results are compared with the experimental results on Amazon EC2 [26] [27], it is observed that there is no significant difference between these results. It is also observed that the effect of buffer size is negligible and depends up on the number of arrivals, number of VM's in operation and requests completion time. The profit model Eq. (1) contains revenue and running cost as given in [26] [27] . The additional value in the Eq. (1) is revenue as per λ and γ. For fixed value of γ there is no change in profit trend but the changes in analytical λ in [26] is compared with experimental results in [27] for obtaining running cost. So the experimental results of the Profit Model exhibit the same trend. Hence there is no necessity to conduct the experimental analysis of the profit model to validate this. If the cloud provider dynamically changes the user request charge γ as per the number of user arrivals which then obviously demands experimental analysis for validation.

VI CONCLUSION

This paper discusses the formulation of profit aware VM provisioning/release for deploying web application on cloud using queueing model. For achieving this formulation and to observe the profit trend analysis for various values of λ and S, two scenarios are considered. They are (i) S and λ varying scenario and (ii) K and λ varying scenario. Observing the numerical results of the scenarios, the S and λ varying scenario exhibits more effect on profit than K and λ varying scenario. Hence the S and λ varying scenario is considered and analysed the profit trend in depth and identified three cases. From these three cases two algorithms were proposed for VM provisioning/release by taking QoS metrics/without QoS metrics in to account. The future scope of this work is to change the unit of charge per request to attain maximum profit for attracting the users by taking QoS metrics and dynamic pricing in to consideration.

REFERENCES

1. Mell P, Grance T. The NIST definition of cloud computing. NIST Special Publication. 2011; 53(6):50
2. I. Foster, I. Yong, Z. Raicu and S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared", Grid Computing Environments Workshop, (2008)
3. Ali, TajEldinSuliman M., and Hany H. Ammar. "Pricing Models for Cloud Computing Services, a Survey." International Journal of Computer Applications Technology and Research 5 (2008).
4. Jin, L. J., and Machiraju, V. A. Analysis on Service Level Agreement of Web Services, (June 2002).
5. M. Al-Roomi, Sh. Al-Ebrahim, S. Buqrais and I. Ahmad, Cloud Computing Pricing Models: A Survey, International Journal of Grid and Distributed Computing Vol.6, No.5, pp.93-106, 2013.
6. Jäätmaa J. Financial aspects of cloud computing business models.2010.
7. Yeo CS, Venugopal S, Chu X, Buyya R. Autonomic metered pricing for a utility computing service. Future Generation Computer Systems. 2010; 26(8):1368-80
8. Samimi P, Patel A. Review of pricing models for grid & cloud computing. In IEEE symposium on computers & informatics 2011 (pp. 634-9). IEEE
9. Mihailescu M, Teo YM. Dynamic resource pricing on federated clouds. In proceedings of the IEEE/ACM international conference on cluster, cloud and grid computing 2010 (pp. 513-7). IEEE Computer Society
10. Amazon EC2 Pricing, http://aws.amazon.com/ec2/pricing/.
11. M. Mihailescu and Y. M. Teo, "Dynamic Resource Pricing on Federated Clouds", Proc. 10th IEEE/ACM Int. Symp. On Cluster. Cloud and Grid Computing, (2010).
12. J. Rohitratana and J. Altmann, "Agent-Based Simulations of the Software Market under Different Pricing Schemes for Software-as-a-Service and Perpetual Software", Economics of Grids, Clouds, Systems, and Services, ser. Lecture Notes in Computer Science, Altmann et al., Eds. Springer Berlin/Heidelberg, pp. 6296. (2010).
13. H. Li, J. Liu and G. Tang, "A Pricing Algorithm for Cloud Computing Resources", Proc. Int. Conference on Network Computing and Inform. Security, (2011).
14. C. S. Yeoa, S. Venugopalb, X. Chua and R. Buyyaa, "Autonomic Metered Pricing for a Utility Computing Service", Future Generation Computer Syst., vol. 26, no. 8, (2010).
15. M. Macias and J. Guitart, A Genetic Model for Pricing in Cloud Computing Markets, Proc. 26th Symp. of Applied Computing, 2011
16. C. Qu, R. N. Calheiros, and R. Buyya, "Auto-scaling Web Applications in Clouds: A Taxonomy and Survey," arXiv preprint arXiv: 1609.09224, 2016
17. Mohammad Sadegh Aslanpoura, Mostafa Ghobaei-Aranib, Adel Nadjaran Toosic," Auto-scaling Web Applications in Clouds: A Cost-Aware Approach", Journal of Network and Computer Applications July 2017
18. Y. Shen, H. Chen, L. Shen, C. Mei, and X. Pu, "Cost-Optimized Resource Provision for Cloud Applications," in High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS), 2014 IEEE Intl Conf on, 2014, pp. 1060-1067
19. S. Lehmann and P. Buxmann, "Pricing Strategies of Software Vendors", Business and Information Systems Engineering, (2009)
20. W. Wang, P. Zhang, T. Lan and V. Aggarwal, "Datacenter Net Profit Optimization with Individual Job Deadlines", Proc. Conference on Inform. Sciences and Systems, (2012).
21. Wang, Deyuan, et al. "Pricing reserved and on demand schemes of cloud computing based on option pricing model." Network Operations and Management Symposium (APNOMS), 2013 15th Asia-Pacific. IEEE, 2013

22. Pal, R. and Hui, P., Economic models for cloud service markets: Pricing and Capacity planning. Theoretical Computer Science 496, 113-124, July. 2013
23. J. Cao, K. Hwang, K. Li, and A. Y. Zomaya, Optimal Multiserver Configuration for Profit Maximization in Cloud Computing, IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 6, pp. 1087-1096, 2013.
24. J. Mei, K. Li, A. Ouyang, and K. Li, A profit maximization scheme with guaranteed quality of service in cloud computing, IEEE Transactions on Computers, vol. 64, no. 11, pp. 30643078, 2015
25. Y. J. Chiang and Y. C. Ouyang, Profit Optimization in SLA-Aware Cloud Services with a Finite Capacity Queuing Model, Mathematical Problems in Engineering, vol. 2014, pp. 01-11, 2014.
26. N.Neelima, B.BasaveswarRao, K.GangadharaRao, K.Chandan , "Performance Analysis of web Application deployment on cloud using M/M/S/K Queueing model" ,International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018) pp. 9485-9492
27. N.Neelima, B.BasaveswarRao, K.GangadharaRao, K.Chandan, "An Experimental Evaluation of Running Cost Analysis for Web Application on Cloud Using Queueing Model", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-3, 2019
28. Aferdita Ibrahimi," Cloud Computing: Pricing Model", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017
29. Muhammad Adeel Javaid," Proposed Pricing Model for Cloud Computing", Computer Science and Information Technology 2(4): 211-218, 2014
30. Donald Gross, John F. Shortle, James M. Thompson, Carl M. Harris "Fundamentals of Queueing Theory", 4th Edition
31. Hao-peng CHEN, Shao-chong L1"A Queueing-based Model for Performance Management on Cloud", https://www.researchgate.net/publication/224219108
32. Khaled Salah et al. (2015). "An Analytical Model for Estimating Cloud Resources of Elastic Services": Springer Journal.
33. Rodrigo N.Calheiros , Rajiv Ranjan, Rajkumar Buyya "Virtual Machine Provisioning Based on Analytical Performance and QoS in Cloud Computing Environments" , International Conference on Parallel Processing 2011