



PERFORMANCE MODELING OF CLOUD E-MARKETPLACE USING DYNAMIC CONTROL MODEL

A. O. Akingbesote, M. O Adigun

Department of Computer Science, University of ZuluLand, P. O. Box X1001, KwaDlangezwa, South Africa

ABSTRACT

One major issue that has been of a great challenge in cloud E-marketplaces is the effective management of resources. To satisfy consumers, some E-market providers have adopted the resource over provisioning model. But literature reveals that resource over provisioning could lead to a largely sub optimal utilization of the hosting environment thereby leading to unnecessary wasteful costs. Also, with server consolidation, about \$45 billion was spent on server management and administrative costs in 2012 alone. Therefore, effective resource management of these Server Machines in terms of cost minimization is imperative. This paper addresses this problem through the use of Dynamic Control System that Changes the number of server machines dynamically based on consumers' waiting time. A comparative analysis of this model is done with fixed servers based on cost. Two issues are addressed in this paper: the first is to determine which of the two models provide optimal solution in terms of server management and the second is the profitability concept of the experiment in the model that produces optimal solution. The results reveal a better cost effective and a competitive Cost-Benefit Ratio (CBR) of DCS over the fixed servers.

Keywords:- E-marketplaces, Dynamic Control System, Cost Benefit Analysis, Optimal Solution, Server Machines.

1. INTRODUCTION

Cloud E-marketplaces can be referred to as the virtual environments for buying and selling of services (Akingbesote, et al 2014a). These services may be on-demand, sport and the reserved. For example, Amazon Elastic Compute Cloud (EC2) offers three different services; the Reserved, Sport and On-Demand (AWS 2015). Also the services on sale may be Infrastructure as a Service (IaaS), Software as a Service (SaaS) and Platform as a service (PaaS)..

One major issue that has been a great challenge in cloud E-marketplaces is the effective management of resources. To satisfy consumers, some E-market providers adopt the resource overprovisioning model (Madhumathi and Ganapathy 2013). But literature reveals that resource over provisioning could lead to a largely sub optimal utilization of the hosting environment thereby leading to unnecessary wasteful costs (Ferretti et al 2010; Pal and Hui 2013). Apart from this, the 2012 IDC Report on power mentioned three problems being faced by cloud E-market providers that include: skyrocketing power consumption and electricity bills, serious

environmental impact, and unexpected power outages. Additionally, with server consolidation, the report in Bailey (20109) shows that about \$45 billion was spent on server management and administrative costs in 2012 (see Fig. 1).

Fig.1 Worldwide IT Spending on Server, Power and Cooling and Management/Administration, 1996 – 2012

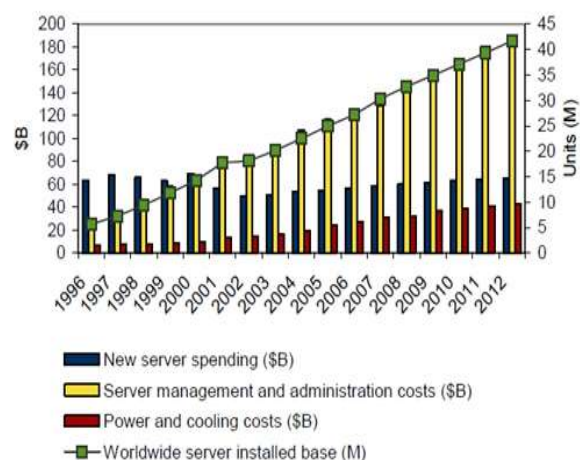


Fig.1 Worldwide IT Spending on Server, Power and Cooling and Management/Administration, 1996 – 2012

As a way of checking over provisioning, some scholars use the optimization modeling method both in single and multi-service offering (Akingbesote et al 2014a; 2014b;2014c). But this method is descriptive or normative, by this model we mean, a model that describes what a real world should be like. This model is sometimes called Optimal Design and Control of queues, Economic or Statics model (Gross and John 1985; Tadjji and Choudhury 2005). It determines the optimal system parameters, for example, the optimal server needed to satisfy consumers. One major issue that was not considered fully is that of server optimization in terms of effective management based on good policy. In this paper, we tackle this issue by focusing on the use of prescriptive model which is an optimal control mechanism that prescribes what the real situation will be (Gross and John 1985; Tadjji and Choudhury 2005) as against descriptive model. Two things we aim to address in

this paper: the first is to determine which of the two models provide optimal solution in terms of server management, while, the second is the profitability concept of the experiment in the model that produces optimal solution. The way of achieve this is to develop a model that will dynamically change based on the input and the history (Hysteresis) of the incoming consumers. The idea is to use a good policy that will control the system economically by optimizing cost. We study three policies: the N (Yadin and Naor19663) , T (Daniel1977) and D (Balachandran 1973). We use the notion N policy because it is analytically easier to deal with [10]. By N policy, we mean a policy that enables the servers to be turned on when the system is busy and turned off when it is empty (Takagi 1991). A comparative of this model is done with classical fixed servers based on cost and

results are extensively discussed.

The remainder of this paper is organized as follows: Section 2 discusses the architecture of the Dynamic Control System (DCS). The required measures of effectiveness are derived in section 3 and the profitability of the system is discussed in section 4 while the experimental set with the results and discussion are discussed in section 5 and 6, respectively. This paper ends with the summary in section 7.

2. ARCHITECTURE OF THE DYNAMIC CONTROL SYSTEM (DCS)

The architecture is shown in Fig. 2. In this Figure, we set up a fixed number of servers for operation (M1, M2, M3) and then provide a limited set of backups as the reservoir (Reservoir Room) with upper limit (S). When the waiting time of service consumer reaches a point say N, then the Dynamic Control System (DCS) transfers the incoming consumers from the dispatcher-In to the reserve center. As soon as the waiting time reduces to N-1, it goes back to the main center. We base our experiment under two conditions. These are:

- ? When the total number of used servers are fixed
- ? When some are fixed with variable service stations as shown in our Fig. 2 (M1..M3 are fixed and M4..M5 varies)

As earlier said, two things we aim to address in this paper: the first is to determine which of the two models provide optimal solution in terms of server management and the second is the profitability concept of the model that produces optimal solution based on the experiment conducted.

To achieve this aim, The following steps are considered:

- i. Determine the Measure of effectiveness
- ii. Subject the two condition to the same test by

Performance Modeling of Cloud E-marketplace Using Dynamic Control Model

- Max
The Server Utilization
Subject to
Waiting time < predetermine Target
- iii. Determine the optimal number of server
- 1

- problem in ii under the two conditions.
- iv. Compare and contrast the result obtained using the Cost-Benefit Analysis (CBR).

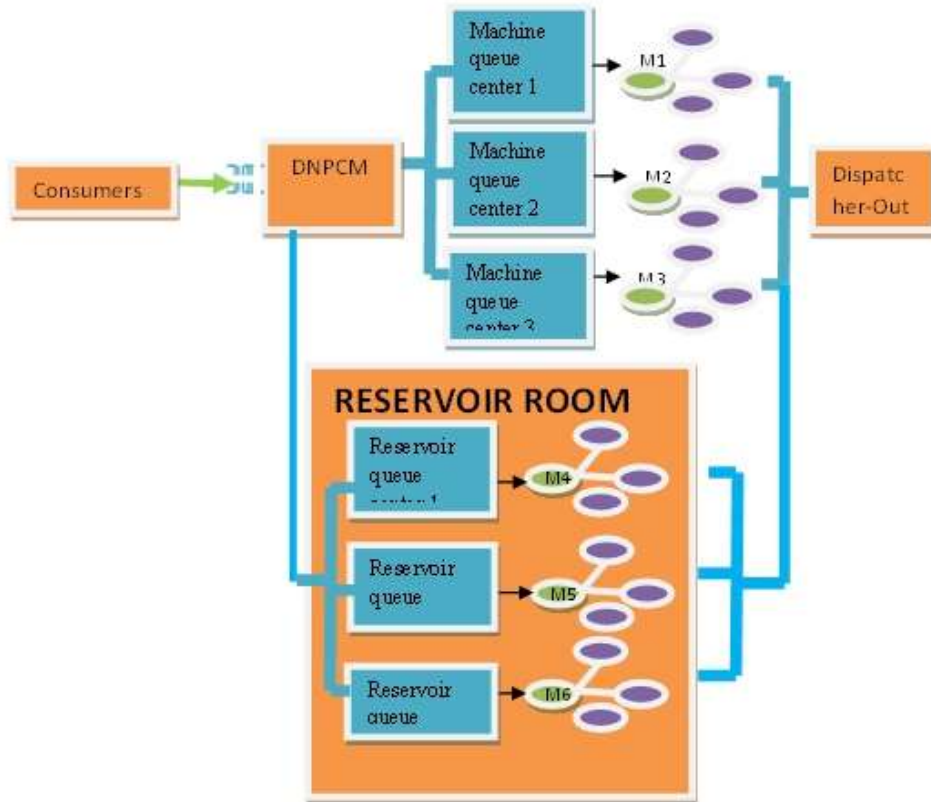


Fig. 2: Architectural Diagram of the Dynamic Control System (DCS)

- v. Find the profitability of using this model over ordinary fixed model based on the experiment conducted using the Cost-Benefit Ratio (CBR).

justification for this is because, it is directly in line with our design but ours is in the contest of cloud E-marketplaces and the issue of cost and benefit was neither discussed nor was the profitability mentioned. Second, it allows hysteresis in the shifting mechanism. Our idea is that, the number of servers is a random variable since it is a function of the queue length. In Moder (1960), the author determines the measure of effectiveness through the following steps (see the full equation in Appendix A):

The analytical solution of this work is based on the use of queuing theory as the proof of concept. We searched various literature (Romani 1957; Moder 1960; Khmelnsky and Gerchak 2002; Li and yang 2000; Pazgal and Radas 2008; Systems 2007; Rate 2015; Yamashiro 1996; Dudin 1981) to identify the most appropriate mathematical concept to support this work. The result of our search enables us to base the mathematical concept on the work in Moder (1960). The

- ⊗ Determine the list of admissible state
- ⊗ Determine the steady state equations
- ⊗ Solve each of the steady state equations

Calculate the measure of effectiveness.

The following parameters are used in this paper

N=queue length

s=number of busy channel

L = mean number of consumers in the system

Lq= mean number of consumers in the queue

$$\delta < S < \infty$$

M= number of (fixed) channel in the

W_q=onventional multiple channel process

Shift up point, i.e the queue length at

which additional channels are

instantaneous $\delta \geq 1$ if $s < S$

Pns = Steady state probability that n consumers are in the system

λ = mean arrival rate

μ = mean service rate per channel

$\rho = \lambda/\mu$ = Utilization factor

$\gamma = \rho/\delta$ Server utilization

Mean waiting time in the queue

Maximum number of manned channels

v = shift down point i.e the minimum queue length when

$$idle_{time} = \sum_{s=0}^{\delta-1} (\delta - s) P_{0s} = P_{00} \sum_{s=0}^{\delta-1} \frac{(\delta-s)\rho^s}{s!} \quad (1)$$

$$Inter_{time} = \sum_{s=\delta}^{\infty} P_{N-s} = \gamma^{N-1} (1-\gamma) P_{00} / (1 - \gamma^{N-v}) \gamma^{v-1} (\rho^v / \delta^v) \quad (2)$$

$$= \frac{\rho^{\delta}}{1 - \gamma^{N-v}} \left[\frac{\rho^{\delta} [\gamma - (v+1)\gamma^{v+1} + v\gamma^{v+2}](1-\gamma^v)}{\delta^v (1-\gamma)^2} + \frac{\rho^{\delta} \{ \gamma^N (N\gamma - \gamma - N) + \gamma^{v+1} (v+1 - v\gamma) \}}{\delta^v (1-\gamma)^2} - \frac{1}{2} \gamma^N [N(N-1) - v(v+1)] \right]$$

$$+ \frac{1}{2} \gamma^{N-1} (1-\gamma) [N(N-1) - v(v-1)] \left\{ \sum_{s=\delta}^{S-1} \frac{\rho^s \rho^{\delta}}{s! \delta^s} \right\}$$

$$+ \frac{\rho^{\delta} \gamma^{N-1} (1-\gamma)}{2S^v (S-\rho)^2 S^{N-2}} \{ S^{N-1} (S-\rho)^2 [(N-1)(N-2) - v(v-1)] - 2S^v \rho^{N-v} [\rho(N-2) - S(N-1)] - 2S^{N-v-1} \rho^{v+v} [Sv - (v-1)\rho] + 2(S^{N-v} - \rho^{N-v}) S^v [S(S(N-1) -$$

$$- \rho(N-2)] \} \quad (3)$$

$$w_q = \frac{L_q}{\lambda}$$

then we

$$\quad (4)$$

$$\left. \begin{array}{l} Max \\ 1 - idle_{time} (server Utilization) \\ \\ subject\ to \\ \\ W_q \leq k \end{array} \right\}$$

Where k is the predetermined waiting time target

4. PROFITABILITY OF THE DCS

One approach to determine the profitability of this model is to use the waiting cost difference between the fixed and the variable operation and see if it is higher than that of waiting cost when additional server

$X_n (E(L) - E(L_q)) > X_{n+1} Prob. (n > N)$ if (6) is true, then it is profitable. To

do this, if the number of consumers under the M/M/n (Fixed) = E(L). Then if a fixed number of server is in operation and the expected number of consumers on the queue is (Lq) under the variable one. if another server is added when $n > N$, then system

becomes profitable to use the second server $X_1 (E(L) - E(L_q)) > X_2 Prob. (n > L)$ (7)

$$= X_1 (E(L) - E(L_q)) > X_2 Prob. (N < n > L) + X_3 Prob. (n > L) \quad (8)$$

where X_n and X_{n+1} are the associated cost for the difference and that of the probability

= 0.4, 0.3, 0.2 respectively. The associated costs (\$) X1, X2, X3, X4, and X5, were set to 4, 4, 2, 2, and 4 respectively. The value of 2 watts was given as the energy consumed when the server is in operation. Each experiment was run for an average of 47 hours and replicated 10 times for accuracy purpose. The results are shown in Fig.3-6 respectively and explanations are given under the results and discussion section.

6. RESULTS AND DISCUSSION

On each of the experiment conducted, a total of 30,000 requests were processed when the arrival time (λ) were set to 0.4, 0.3 and 0.2 as shown in Fig. 3 and 4. These two figures show the number of consumers that were processed by each machine under DCS (Fig.3) and the fixed classical model (Fig. 4). Under the DCS, we observed that as the service time reduces, the number of servers increases in DCS. In the first experiment, when $\lambda = 0.4$ the maximum number out of the five manned servers that were used were two as against five that was used in the Classical Fixed Model (CFM). We recorded a gain of three server machines which were used to generate our server benefit based on equation 11. Another gain of two servers were recorded in the experiment when $\lambda = 0.3$ while one was recorded when $\lambda = 0.2$. The reason for the gain in server machines used under the DCS is explained using Fig. 5 and 6. These figures represent the server utilization with the number of machine used. In these two figures, In each of the experiment conducted, the instantaneous utilization of the DCS is higher than the classical fixed model. For example, when the service time was 0.4, the instantaneous utilization of M1 under DCS was about 0.739 as against the classical fixed one which was about 0.304. Also, when the service time was 0.2 in the second experiment, the instantaneous server utilization was about 0.990 by server M1 in

DCS as against 0.603 in CFM. The maximization usage of this server under the DCS is as a result of the dynamic control mechanism introduced to control the operation and this accounted for the gain in server machine used. The DCS has a better performance in terms of server utilization and optimal server provisioning as against the classical fixed model that over satisfies the delay requirement.

In regard to the profitability concept of this model, we recorded the waiting time, idle time, interrupt time and the number of used servers. Each server used was allotted 2 watts as shown in Table 1. The results obtained from our simulation are shown in Table 1. In this Table, as the service time increases in each experiment, so does the waiting time, while the server utilization increases under both conditions. Unlike the waiting time, we recorded zero for interrupt time in the first experiment under the DCS and in all the fixed ones because only two fixed machines worked under the DCS in the first experiment while all the machines were in operation under the fixed model thereby recorded no interrupt period.

The output of Table 1 was used to formulate our Cost-Benefit Table (see Table 2). Using Equations 10, 11 and 12. In this Table (Table 2) the total cost increases while the benefit decreases as we increased the service time as shown under experiment 1, 2, 3. The reason for this is due to the high gain of server recorded in experiment 1 when only two servers were in operation as against the five used in the fixed operation. This also accounted for the high CBR recorded in experiment 1.

One noticeable thing which requires questioning in Table 1 is the low server utilization in experiment 2 compared to experiment 1. The reason is that when $\lambda = 0.3$, though the server utilization in the

(λ)

that the number of consumer is greater than N respectively.

If again we have another server added when $n > L$ with associated cost then it becomes profitable to use the third server when

Our idea is slightly different from these because, the cost in a typical E-marketplace might not likely be based on waiting cost alone. In addition, the benefit accrued to this might also not likely be based on one parameter alone. In our own proposed solution, the DCS becomes profitable if the opportunity cost for adding additional server is less than the benefit incurred when additional number of servers is/are added. We use the Cost-Benefit Analysis (CBA) as our solution approach (Weber 2008).

In our model, we base the opportunity cost on the following parameters:

- ⌘ The average waiting time cost (W_{qave}) incurred for staying longer in the DCS model than the fixed one.
- ⌘ Interrupt Time cost ($Inter_{time\ ave}$) that occurs when additional servers are added.

The benefits are based on the following parameters:

$$Total_{cost} = X_1 W_{qave} + X_2 Inter_{time\ ave} \quad (9)$$

$$= X_1 [W_{qDCS} - W_{qfixed}]$$

$$+ X_2 \left(\left(\sum_{i=1}^r inter_{time} \right) / r \right) \quad (10)$$

$$Total_{benefit} = X_3 [serverused_{fixed} - serverused_{DCS}]$$

$$+ X_4 [Energyused_{fixed} - Energyused_{DCS}]$$

$$+ X_5 [Utilisation_{DCS} - Utilisation_{fixed}] \quad (11)$$

Our DCS model becomes

- Marginal iff $Total_{cost} = Total_{benefit}$
- Profitable iff $Total_{cost} < Total_{benefit}$ and
- Non profitable iff $Total_{cost} > Total_{benefit}$

The Cost-Benefit ratio (CBR) is given as

$$CBR = \frac{Total_{benefit}}{Total_{cost}} \quad (12)$$

Given n number of experiments conducted and the CBR for each experiment is E_1 , E_2 and E_3 , then the optimal profit occurs in the experiment where CBR has the highest value.

5. EXPERIMENTAL SETUP

This experiment is conducted with the assumption that the arrival process is a random process and the service time is exponential. Also, the cost of using each server is the same. To achieve this, we set up our experiment under two conditions:

- i. When the total number of used servers are fixed.
- ii. Under the Dynamic Control system (DCS) in which some are fixed and some are varied as shown in our Fig. 2 (M1, M2, M3 as fixed and M4, M5 and varies).

Under each of the above conditions, three experiments are conducted. In all the experiments we used 5 server machines for both conditions. The service time was homogenous and set to 0.5 seconds. In each experiment, the fixed and the variable (1/λ) were subjected to the same waiting time condition. ($W \leq K = 0.0006$ sec). The arrival is a random process and the service rate is exponential process. Under the first condition, all used servers are fixed (5 servers) and under the second condition, two are fixed and we vary the other ones (3) and the shift point N is set to 5.

Arena simulator is used as our simulator. Under the two conditions, our arrival time

Performance Modeling of Cloud E-marketplace Using Dynamic Control Model

fixed machine under DCS (See Fig. 5) were higher than that of experiment 1, but the low server utilization of the third machine in experiment 2 (M3 indicated by red colour in Fig. 5) accounted for the average drop in server utilization of experiment 2 computation. Even with this drop, a gain of 2 server machines was recorded.

All the three experiments were observed to be profitable due to the Cost-Benefit Ratios that are greater than unity (183.6514: 30.54533: 8.881209) in all the output of all the experiments. The optimal profitability was recorded in experiment 1 with the value of 183.6514. The reason for this is the high gain of the unused servers and the server utilization in the first experiment. For example, 2 servers were used under the

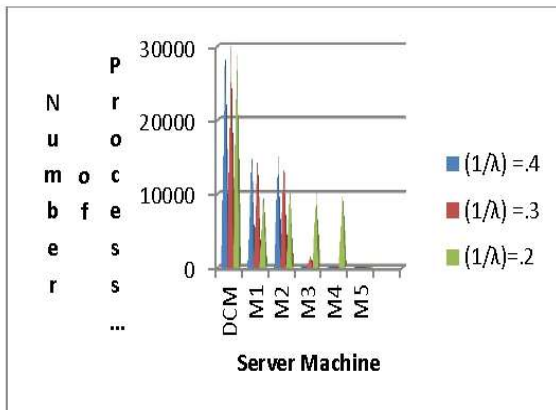


Fig. 3: Number of Processed Consumers Under DCS

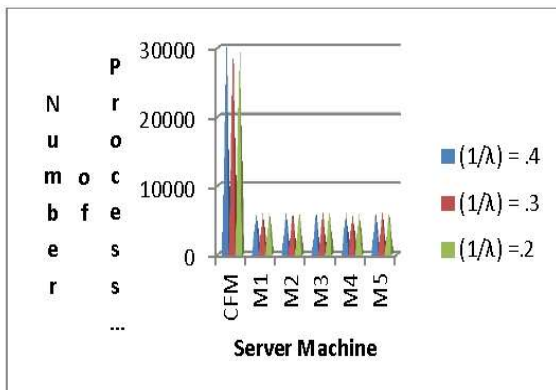


Fig. 4: Number of Processed Consumers under the Classical Fixed Model.

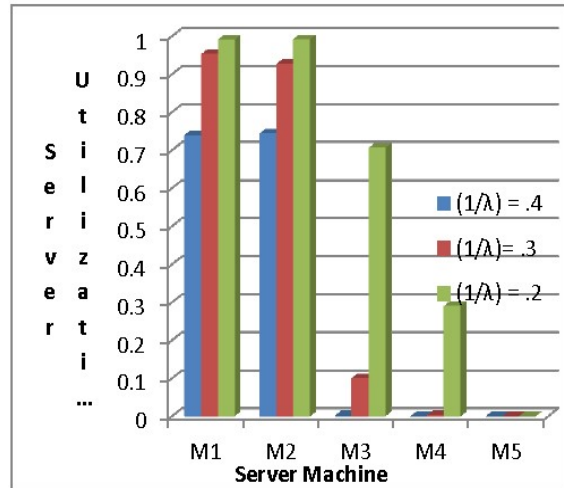


Fig. 5: Server Utilization Under DCS

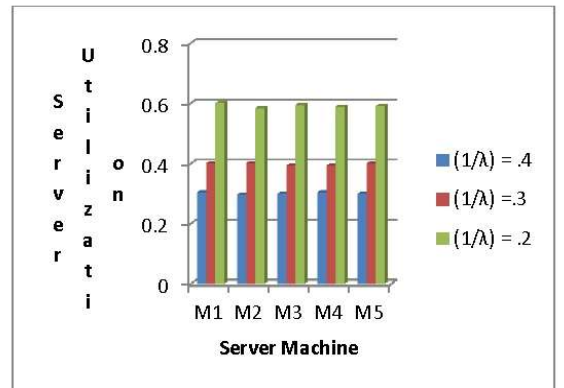


Fig. 6: Server Utilization Under the Classical Fixed Model (CFM).

Table 1: Detail Results Obtained Based on the used Parameters under DCS and the Classical Fixed Methods

Parameter	DCS			Fixed		
	Experiment 1	Experiment 2	Experiment 3	Experiment 1	Experiment 2	Experiment 3
Waiting time (sec)	0.000547	0.001315	0.002698	0.000303	0.000357	0.000544
Ave. Server Utilization	0.742531	0.661123	0.991867	0.3309	0.3575	0.5915
Interrupt Time (min)	0.00	.00000004	0.0000072	0.0	0.0	0.0
Servers used	2	3	4	5	5	5
Power consumed (Watt)	4	6	8	10	10	10

Table 2: Cost-Benefit Analysis of DCS and Classical Methods

Cost -Benefit			
	Experiment 1	Experiment 2	Experiment 3
Total Cost (\$)	0.058625	0.230899	0.518563
Total Benefit (\$)	10.76652	7.052892	4.605468
Cost-Benefit Ratio (\$)	183.6514	30.54533	8.881209

CONCLUSION

In this paper, we have investigated the N-policy in the cloud E-marketplaces using a Dynamic Control System (DCS). Two issues we have addressed in this paper: the first is to determine which of the two models provide optimal solution in terms of server management and the second is the profitability concept of the experiment in the model that produces optimal solution.

This was done by manning some servers and varying some. On the first one, a comparative study of DCS model with the classical M/M/s was undertaken. Our results reveal a better performance in terms of server utilization and optimal server provisioning as against the classical model that over satisfies the delay requirement. On the second one, a systematic Cost-Benefit Analysis (CBA) was developed to determine the profitability concept. The profitability is a function of the cost and other parameters used. Since the CBA indicated that the benefit accrued is greater

than the cost, which is greater than unity in all the experiments conducted, hence it is concluded that the DCS is profitable. Moreover, our Cost-Benefit Ratio (CBR) showed the experiment that produced the optimal profit. This is experiment 1 and it has the value of 183.6514 as indicated by the yellow colour in Table.2. This model is hoped to further guard against server under provisioning and over provisioning in the cloud E-marketplaces. It will be a good prescriptive mechanism for effective management of the Cloud E-marketplaces.

ACKNOWLEDGMENT

This work is based on the research supported in part by the National Research Foundation of south Africa-grant uid: tp11062500001 (2012-2015). The authors also acknowledge funds received from industry partners: Telkom SA Ltd, Huawei Technologies SA (PTY) Ltd and Dynatech Information Systems, South Africa in support of this research.

Performance Modeling of Cloud E-marketplace Using Dynamic Control Model

Appendix A – Derivation of measures of effectiveness

A.1 Determination the list of admissible state

The admissible state are written below

$$0 \leq s \leq \delta - 1 \quad \text{then } n = 0 \quad (\text{A. 1})$$

$$s = \delta \quad \text{then } 0 \leq n \leq N - 1 \quad (\text{A. 2})$$

$$\delta + 1 \leq s \leq S - 1 \quad \text{then } v \leq n \leq N - 1 \quad (\text{A. 3})$$

$$s = S \quad \text{then } v \leq n \quad (\text{A. 4})$$

A.2 Derivation of the steady state equation

Knowing that the mean transition rate of inflow of one subset is the same as the mean transition rate out of that subset, then the steady state equations are given as:

$$\lambda P_{0,s} = \mu(s+1)P_{0,s+1} \quad (\text{A. 5})$$

$$\lambda P_{n,\delta} = \delta \mu P_{n+1,\delta} \quad (0 \leq n \leq v-1, s = \delta)$$

$$\lambda P_{n,\delta} = \delta \mu P_{n+1,\delta} + \lambda P_{N-1,\delta} \quad (v \leq n \leq N-2, s = \delta) \quad (\text{A. 7})$$

$$\lambda P_{n,s} + s \mu P_{v,s} = s \mu P_{n+1,s} + \lambda P_{N-1,s} \quad (\text{A. 8})$$

$$s \mu P_{v,s} = \lambda P_{N-1,s-1} \quad (\delta + 1 \leq s \leq S) \quad (\text{A. 9})$$

$$\lambda P_{n,s} + S \mu P_{v,s} = S \mu P_{n+1,s} \quad (v \leq n \leq N-2, \delta + s = S) \quad (\text{A. 10})$$

$$\lambda P_{n,s} = S \mu P_{n+1,s} \quad (N-1 \leq n) \quad (\text{A. 11})$$

A.3 Derivation of the steady state probability

A general approach is used to determine the steady state probability.

Writing

$$P_{0,s+1} = \left[\frac{\rho}{s+1} \right] P_{0,s}$$

and let $s=0, s=1, \dots$ can be expressed as a function of P_{00} as shown in equation A.12. then

$\sum_1^k P_{n,s} = 1$ where k is the total number of the admissible state.

The summary to the solutions of equation A.5 to A.11 is given below

$$P_{0,s} = \left(\frac{\rho^s}{s'} \right) P_{00} \quad (n = 0, 0 \leq s \leq \delta) \quad (\text{A. 12})$$

$$P_{n,\delta} = \gamma^n \left(\frac{\rho^\delta}{\delta'} \right) P_{00} \quad (0 \leq n \leq v, s = \delta) \quad (\text{A. 13})$$

$$P_{n,\delta} = \left(\frac{\rho^\delta}{\delta'} \right) \left[\frac{(\gamma^n - \gamma^N)}{(1 - \gamma^{N-v})} \right] P_{00} \quad (v \leq n \leq N-1, s = \delta) \quad (\text{A. 14})$$

$$P_{n,s} = P_{v,s} \quad (v \leq n \leq N-1, \delta + 1 \leq s \leq S-1) \quad (\text{A. 15})$$

$$P_{n,s} = \left(\frac{\rho^\delta}{s'} \right) \left[\left(\frac{\gamma^{N-1} - \gamma^N}{1 - \gamma^{N-v}} \right) \right] \quad (v \leq n \leq N-1, \delta + 1 \leq s \leq S-1) \quad (\text{A. 16})$$

$$P_{n,s} = \rho^s \left[\left(\frac{(S^{n-v+1} - \rho^{n-v+1})(\gamma^{N-1} - \gamma^N)}{S^{n-v} S' (S - \rho)(1 - \gamma^{N-v})} \right) \right] P_{00} \quad (v \leq n \leq N-1, s = S) \quad (\text{A. 17})$$

REFERENCES

- [1] A. O. Akingbesote, M. O. Adigun, S. S. Xu, and E. Jembere, "Performance Evaluation of Cloud E-Marketplaces using Non Preemptive Queuing Model," in **WCST conferece**, 2014, pp. 66–71.
- [2] "AWS | Amazon Elastic Compute Cloud (EC2) - Scalable Cloud Hosting." [Online]. Available: <http://aws.amazon.com/ec2/>. [Accessed: 23-Apr-2015]."
- [3] C. Madhumathi and G. Ganapathy, "An Academic Cloud Framework for Adapting e-Learning in Universities," **Int. J. Adv. Res. Comput. Commun. Eng.**, vol. Vol. 2, Is, 2013.
- [4] S. Ferretti, V. Ghini, F. Panzieri, M. Pellegrini, and E. Turrini, "QoS–Aware Clouds," **2010 IEEE 3rd Int. Conf. Cloud Comput.**, pp. 321–328, Jul. 2010.
- [5] R. Pal and P. Hui, "Economic models for cloud service markets: Pricing and Capacity planning," **Theor. Comput. Sci.**, vol. 496, pp. 113–124, Jul. 2013.
- [6] M. Bailey, "The Economics of Virtualization : Moving Toward an Application-Based Cost Model," 2009.
- [7] A. O. Akingbesote, M. O. Adigun, J. Oladosu, E. Jembere, I. Kaseeram, S. Africa, and S. Africa, "Modeling the Cloud e-Marketplaces for Cost Minimization Using Queuing Model," **Aust. J. Basic Appl. Sci.**, vol. 8, no. 4, pp. 59–67, 2014.
- [8] A. . Akingbesote, M. . Adigun, E. Jembere, and J. Oladosu, "Modeling the Cloud e-Marketplaces for Cost Minimization Using Queuing Mode," **Aust. J. Basic & Appl. S c i , 8 (4) : , 2 0 1 4 . [O n l i n e] . A v a i l a b l e : <http://ajbasweb.com/old/ajbas/2014/Special/59-67.pdf>. [Accessed: 16-Apr-2015].**
- [9] D. Gross and F. John, "Fundamentals of Queueing Theory," **Wiley Series in Probability and Mathematical Statistics 2nd Edition**, 1985. [Online]. Available: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118211642.html>. [Accessed: 02-Oct-2014].
- [10] L. Tadj and G. Choudhury, "Optimal design and control of queues," **Top**, vol. 13, no. 2, pp. 359–412, 2005.
- [11] M. Yadin and P. Naor, "Queueing Systems with a Removable Service Station†," **J. Oper. Res. Soc.**, vol. 14, no. 4, pp. 393–405, 1963.
- [12] Daniel P. Heyman, "The T-Policy for the M/G/1 Queue," **Management Science: I N F O R M S , 1 9 7 7 . [O n l i n e] . A v a i l a b l e : <http://pubsonline.informs.org/doi/pdf/10.1287/mnsc.23.7.775>. [Accessed: 10-Jun-2015].**
- [13] K. R. Balachandran, "Control Policies for a Single Server System," **Manage. Sci.**, vol. 19, no. 9, pp. 1013–1018, 1973.
- [14] H. Takagi, "Queueing analysis A foundation of Performance Evaluation Volume 1: Vacation and priority systems," **Volume 1: Vacation and priority systems**, 1991. [Online]. Available: <http://libra.msra.cn/Publication/2404848/queueing-analysis-a-foundation-of-performance-evaluation-volume-1-vacation-and-priority-systems>. [Accessed: 10-Jun-2015].
- [15] J. Romani, "Un modelo de la teoria de colas con numero variable de canales," **Trab. Estad.**, vol. 8, no. 3, pp. 175–189, Oct. 1957.
- [16] J. J. Moder, "Queueing with fixed and variable channels," pp. 218–232, 1960.

- [17] E. Khmelnitsky and Y. Gerchak, "Optimal control approach to production systems with inventory-level-dependent demand," ... **Control. IEEE Trans.**, pp. 1–12, 2002.
- [18] H. Li and T. Yang, "Queues with a variable number of servers," **Eur. J. Oper. Res.**, vol. 124, no. 3, pp. 615–628, 2000.
- [19] A. I. Pazgal and S. Radas, "Comparison of customer balking and renegeing behavior to queueing theory predictions: An experimental study," **Comput. Oper. Res.**, vol. 35, no. 8, pp. 2537–2548, 2008.
- [20] M. M. Systems, "Queueing system with variable server number," no. 4, pp. 63–65, 2007.
- [21] A. Rate, "OPTIMAL CONTROL OF A SERVICE FACILITY WITH," vol. 18, no. 9, pp. 560–566, 2015.
- [22] M. Yamashiro, "A system where the number of servers changes depending on the queue length," **Microelectron. Reliab.**, vol. 36, no. 3, pp. 389–391, 1996.
- [23] A. i A. N. Dudin, "Optimal assignment of the rate for service of customers in a multilinear two-rate service system," **Telemekh.**, no. 11, 1981. [Online]. Available: http://www.mathnet.ru/php/archive.phtml?ws how=paper&jrnid=at&paperid=6044&option_lang=eng. [Accessed: 10-Jun-2015].
- [24] S. S. Xu, "Queue-dependent servers," **J. Eng. Math.**, vol. 7, no. 2, pp. 123–126, Apr. 1973.
- [25] R. L. Garg and P. Singh, "Queue-dependent servers queueing system," **Microelectron. Reliab.**, vol. 33, no. 15, pp. 2289–2295, Dec. 1993.
- [26] T.A. Weber, **Price Theory in Economics**. 2008.