Multi-Criterion VM Selection Policy for Workload Consolidation in Cloud Computing

Nimisha Patel
Research Scholar, Rai University, Ahmedabad
Associate Professor, Computer Engineering Department, Sankalchand Patel College of Engineering, Visnagar
Gujarat, India
nimishaa_25@yahoo.co.in

Dr. Hiren Patel
Professor, Computer Engineering Department, LDRP Institute of Technology & Research,
Gandhinagar-Gujarat, India
hbpatel1976@gmail.com

***Abstract*— As a part of workload consolidation in Cloud datacenter, after detecting overloaded hosts, it is required to select the VMs to be migrated such that after migration, the overloaded hosts become normal. Selection of VMs to be migrated is a crucial decision as it affects the performance of both the types of hosts (i.e. the overloaded hosts from which the VMs are selected and the hosts on which the selected VM are placed). This in turn impinges on the performance of overall datacenter. We propose a policy for selection of VMs to be migrated from overloaded hosts based on three criterions. Select VMs in such as way that (i) after migrating the VMs, utilization of overloaded hosts falls below upper threshold value and remain nearest to it (*MaxUtil*) (ii) the number of VMs remains minimal (*MinTotalMigr*) and (iii) resultant migration time stays minimum (*MinMigrTime*). All these three parameters are significant in the process of workload consolidation. Importance of individual parameter over the other is left to individual user's requirement. In this research, we have proposed a method to optimally select a best VM combination such that all these three parameters are satisfied.**

Keywords-workload consolidation; VM Selection; overloaded host

I. INTRODUCTION

Over the last decade, economy has moved from traditional paper-based to digitization. The era of digitization requires numerous data centres to facilitate its primary need for data processing, storage, and communications. Virtually found in all sectors (e.g. financial sectors, academics, government organization, industry etc), data centres are now inevitable part of one's life directly or indirectly. Over the last few years, Cloud has adhered to this list and grown significantly. All major drifts in information communication technology (ICT), viz. Cloud Computing, Big Data and Mobile Computing require backup of prevailing computing infrastructure support. Offering a new computational model viz. "Computing as a Service", due to its abundant benefits, Cloud computing has engrossed the concentration of academicians, researchers and of course, the end users.

This paper aspires to attend the broad issue of energy consumption by Cloud data centre and subsequent carbon discharge. To understand the problem, let us take an example of real data centre [1]. Consider a data centre containing 1000 servers the average electrical power entering the system is 1 MW, of which 450 kW are used to operate the servers, 250 kW for the power distribution systems and the UPS devices, and 300 kW for the cooling system. Let us further assume that the average utilization of server resources (CPU, RAM etc.) is 30%. We may compute the Power Usage Effectiveness (PUE) using the equation defined by [2], (power consumption for the power distribution system + power consumption for the cooling system + power consumption for operating the servers) / (power consumption for operating the servers), PUE would come to 2.22. It is evident from this figures that if we employ relatively better and efficient policy for server workload, we may end up with accommodating more number of existing applications on less number of servers. That is, for example, if we can increase the average utilization of server resources from 30% to 60%, PUE may be improved.

There are quite a few ways for optimizing the server usage and to improve computational efficiency of the data centre. Workload consolidation, offered under the umbrella of virtualization technology, is one of such important mechanisms. Figure 1 shows an example to understand workload consolidation.

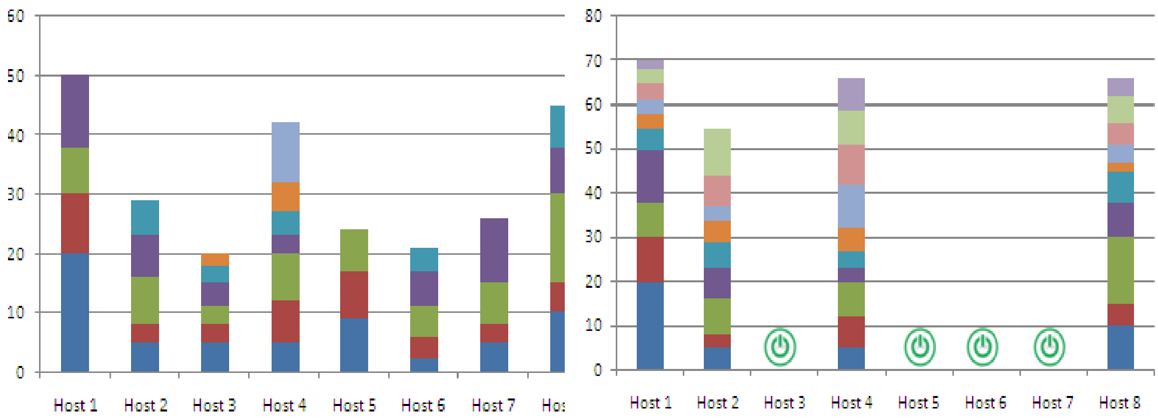

Figure 1. Workload Consolidation. Before consolidation (left) and after consolidation (right)

Figure 1 (Left) shows the status of workload before consolidation. It is seen that all eight hosts are utilized with utilization varying from 20% to 50%. Using the workload consolidation, we may shift few workloads from one host to another in such a way that the target host does not get overloaded. Figure 1 (Right) shows the status of workload after consolidation. It depicts that we could turn four hosts (3, 5, 6 and 7) into power saving mode leaving four hosts (1, 2, 4 and 8) active with utilization ranging from 55% to 70%.

Overall organization of this paper is as follows. Section II illustrates the literature survey in the domain of identifying underloaded server for workload consolidation. Section III depicts our proposal of incorporating the multiple criterions. Section IV illustrates the entire setup of experimentation with different testbed and subsequent results and their discussions. At the end, in section V we conclude our research work followed by list of reference mentioned in Section VI.

II. RELATED WORK

Beloglazov and Buyya [3] categorize the process of VM allocation into two parts. First, selecting the VMs to be migrated and then, placing the selected VMs on hosts. For selecting VMs to be migrated, there are two types of hosts to be identified viz. over-utilized and under-utilized based on some threshold values. These values may be compute either statically or dynamically. In case of under-utilized hosts, it is recommended to select all VMs for migration, and subsequently turn the host to power saving mode. But for the case of over-utilized hosts, few VMs are be selected in such a way that the host's utilization falls below the upper threshold value, for which authors have proposed techniques such as Minimization of Migration (MM), Highest Potential Growth (HPG) and Random Choice (RC). Further, the authors have proposed Dynamic Threshold (DT) method for deciding the upper threshold dynamically and Modified Best Fit Decreasing (MBFD) method for VM Placement. Continuing the work, the same authors Beloglazov and Buyya [4] propose heuristics for various phases of workload consolidation. For Overloaded host detection, authors propose to use statistical methods such as Median Absolute Deviation (MAD), Interquartile Range (IQR), Local Regression (LR) and Robust Local Regression (LRR). For VM selections, policies such as Minimum Migration Time (MMT), Random Choice (RC) and Maximum Correlation (MC) have been proposed. Power Aware Best Fit Decreasing (PABFD) has been proposed for VM placement. For underloaded host detection, authors propose a plain approach where the host with minimum utilization (compared to other hosts) is selected and all the VMs from this host are to be place on other hosts keeping them not overloaded. This process is iteratively repeated. To the best of our understanding, there has not been significant contribution (except mentioned above) in the domain of VM selection, though few other researchers have used the methods proposed above such as Horri,Mozafari and Dastghaibyfard [5], Wadhwa and Verma [6] and Huang, Wu and Moh [7].

We could learn from the literature survey that there are multiple criterions for VM selection. These criterions are (a) keeping the active hosts as utilized as possible while not allowing them to exceed upper threshold (b) keeping the number of VMs to be migrated as minimum as possible and (c) keeping the migration as low as possible. Significance of one criterion over another depends on many factors such as quality of service, service level agreement, and energy efficiency etc. To the best of our understanding none of the researchers have recommend to propose a technique which makes use of these multiple criterion.

Hence, to address the issues and to incorporate these criterions in proposed model, few questions need to be answered. Under which situation the VMs are to be migrated? Which VMs are to be migrated? Where to place the VMs selected for migration? While answering these questions, one need to ensure SLA and other overheads such as migration cost, number of migration, performance etc. Therefore, in this research we aim to propose multi-criterion VM selection policy for workload consolidation in Cloud computing.

III. OUR PROPOSAL

Selection of VM from an over-utilized host is a challenging task. It is exigent because it requires many factors to be taken into considerations such as (i) once selected VM is migrated, the process should not be undone due to any reason (ii) once the selected VM is migrated, the source host should tend to move towards normal state from over-utilized status (iii) the selected VM should not turn the targeted host into over-utilized state (iv) out of multiple selection of VMs, an optimal combination of VMs should be selected in such a way that number of VMs remains minimal and/or migration time stays least (v) all the active hosts should work at their peak ability (defined by upper threshold).

We propose a multi-criteria decision making (MCDM) policy based on three criterion viz. *MaxUtil*, *MinTotalMigr* and *MinMigrTime* as shown in "(1)". The criteria *MaxUtil* illustrates the difference between THRESH_UP and host utilization after migration. In other words, the utilization should reach nearer and below to THRESH_UP after VM migration for both the source and destination host. The criteria *MinTotalMigr* exemplifies that total number of VMs to be migrated should remain as minimal as possible. And the criteria *MinMigrTime* shows that the time consumed by the VMs for migration should remain as least as possible to improve the performance of overall system. Hence, based on all these aspects, we propose a policy for VM selection to calculate selection index (SI) as stated in "(1)".

$$SI = (\alpha * MaxUtil) + (\beta * MinTotalMigr) + (\gamma * MinMigrTime) \qquad \text{where } \alpha + \beta + \gamma = 1 \text{ and } 0 \le \alpha, \beta, \gamma \le 1 \tag{1}$$

In "(1)", α, β and γ are the constant weight ranging between 0 to 1 and the value of the same can be selected by the user based on its requirement in terms of Service Level Agreement (SLA). It is important to define a method to select of weights for α, β and γ. For the purpose of experimentation, the equation 1 cannot be used in its primitive form due to two reasons viz. (i) the units of *MaxUtil* (percentage), *MinTotalMigr*(number) and *MinMigrTime* (µS or nS) are not same and (ii) the value of these units is inversely proportional to its significance, for example, smaller the value of *MinTotalMigr* higher its significance. So, we need to convert the equation on a uniform scale while considering the proportionality of the factors. Hence, we redefine the "(1)" as mentioned beneath in "(2)".

$$SI = \alpha * \left(\frac{MaxUtil_{HIGH} - MaxUtil_{CUR}}{MaxUtil_{HIGH}} \right) + \beta * \left(\frac{MinTotalMigr_{HIGH} - MinTotalMigr_{CUR}}{MinTotalMigr_{HIGH}} \right) + \gamma * \left(\frac{MinMigrTime_{HIGH} - MinMigrTime_{CUR}}{MinMigrTime_{HIGH}} \right) \tag{2}$$

where, HIGHEST represents the maximum value in a given set for a particular variable and CURRENT represent the value being used from the set for experimentation.

In MCDM, the weight determination methods can be either compensatory or outrankable [8]. Compensatory method includes Analytical Hierarchy Process (AHP), Fuzzy Multi-Criteria Decision Making, Process (FDM) etc. Out-ranking method includes ELimination and Choice Expressing Reality (ELECTRE), Preference Ranking, Organization Method for Enrichment of Evaluations (PROMETHUS). The discussion on these methods is beyond the scope of this literature.

In our next section of experimentation and result, we have calculated all three criterions taking sample dataset.

IV. EXPERIMENTATION AND RESULTS

Table 1 illustrates the specifications of the both types of hosts. We have taken 4 hosts into consideration with other details including number of VMs per host and their requirements as mentioned in Table 2.

TABLE 1. SPECIFICATIONS OF HOSTS

	Name	MIPS	RAM (MB)	Bandwidth (Gb per sec)	Core / Processing Elements
Type 1	HpProLiantMl110G4Xeon3040	1860	4096	1	2
Type 2	HpProLiantMl110G5Xeon3075	2660	4096	1	2

TABLE 2. VM REQUEST SPECIFICATIONS

Host ID	Host Type (As per Table 1)	VM ID	VM RAM Requirement (MB)	VM Capacity (MIPS)	VM Core	VM Utilization Required (%)	VM MIPS Required
1	Type 2	101	1800	2500	1	30	750
		102	1600	2000	1	25	500
		103	1200	1000	1	25	250
		104	600	500	1	30	150
		105	1200	1000	1	20	200
		106	1600	2000	1	35	700
		107	1200	1000	1	35	350
		108	1800	2500	1	25	625
		109	600	500	1	40	200
2	Type 1	201	1800	2500	1	24	600
		202	2500	2000	1	21	420
		203	1200	1000	1	18	180
		204	600	500	1	17	85
		205	1200	1000	1	13	130
		206	1600	2000	1	7	140
		207	1200	1000	1	35	350
		208	1800	2500	1	30	750
		209	600	500	1	22	110
		210	1600	2000	1	30	600
3	Type 2	301	1800	2500	1	50	1250
		302	1600	2000	1	60	1200
		303	1200	1000	1	15	150
		304	600	500	1	18	90
		305	1200	1000	1	17	170
		306	1600	2000	1	12	240
		307	1200	1000	1	26	260
		308	1800	2500	1	28	700
		309	600	500	1	31	155
4	Type 1	401	1800	2500	1	26	650
		402	1600	2000	1	39	780
		403	1200	1000	1	29	290
		404	600	500	1	21	105
		405	1200	1000	1	14	150
		406	1600	2000	1	20	400
		407	1200	1000	1	10	100
		408	1800	2500	1	27	675
		409	600	500	1	27	135
		410	1600	2000	1	17	340

We have conducted series of experimentation to generate results as per criterion mentioned in the abstract. For dynamic value setup for THRESH_UP, utilization of HOST ID 1 and 3 remained beneath THRESH_UP. But, HOST ID 2 and 4 were over utilized. So, need arise to move few VMs from these hosts (#2 and #4), such that one or more of the three criterions mentioned in abstract are satisfied.

Table 3 shows the summarized result for HOST ID 2. We have computed the result for exhaustive available options using the combinatorial formula mentioned beneath where where *n* is total number of VMs on a host and *r* varies from 1 to n.

$$nCr = \frac{n!}{r!*(n-r)!}$$

Due to lack of space, we could not accommodate all the results in the table, but have included the best result amongst the available ones from the range. For example, Row #1, may have 10 different results (one for each of the 10 VMs), but it contains only the result of VM# 202 as it gives minimum utilization difference after migration (*MaxUtil*) (i.e. THRESH_UP minus host utilization after VM migration) while considering the fact that the resultant utilization falls below the upper threshold. Likewise, all other rows correspond to different number of combinations of the VMs. As mentioned in abstract, based on individual user's requirement, one may go for optimal option for migrating VM based on *MaxUtil*, *MinTotalMigr* and *MinMigrTime*. For instance, if the user's requirement is to minimize number of VM migrations (*MinTotalMigr*), then in that case, migrating the VM# 202 would be the best optimal solution. On the contrary, if the user's requirement is to keep the resultant host utilization nearer and beneath the THRESH_UP (*MaxUtil*), then in that case, third row would be the best possible option, i.e. migrating three VMs (viz. VM ID #203, #204 and #205) would result into minimum difference between THRESH_UP and host utilization after VM migration, amongst all the nCr combinations.

TABLE 3. NUMBER OF VMS TO BE MIGRATED AND DIFFERENCE OF UTILIZATION AFTER MIGRATION

Number of VMs to Migrate	Actual VMS to Migrate (VM Ids)	Utilization Difference After Migration
1	202	0.00833
2	204,207	0.01237
3	**203,204,205**	**0.00161**
4	204,205,206,209	0.02043
5	203,204,205,206,209	0.06882
6	203,204,205,206,207,209	0.16290
7	202,203,204,205,206,207,209	0.27581
8	201,202,203,204,205,206,207,209	0.43710
9	201,202,203,204,205,206,207,209,210	0.59839
10	201,202,203,204,205,206,207,208,209,210	0.80000

Figure 2 depicts the difference between the resultant host utilization (after migration) with upper threshold (*MaxUtil*). Lesser the value of this difference, optimal the solution. As we can see from graph, the combination with number of VM=3 is optimal. Experiment results show that a combination of three VMs {203,204,205} yields to closest value to upper threshold with the difference 0.001612903225806539. Whereas, the combination of one VM {202} yields second closest value to upper threshold with the difference 0.008333333333333415 (*MinTotalMigr*). Hence, it is recommended to select the combination with number of VM=3, as far as host utilization after migration is concerned.

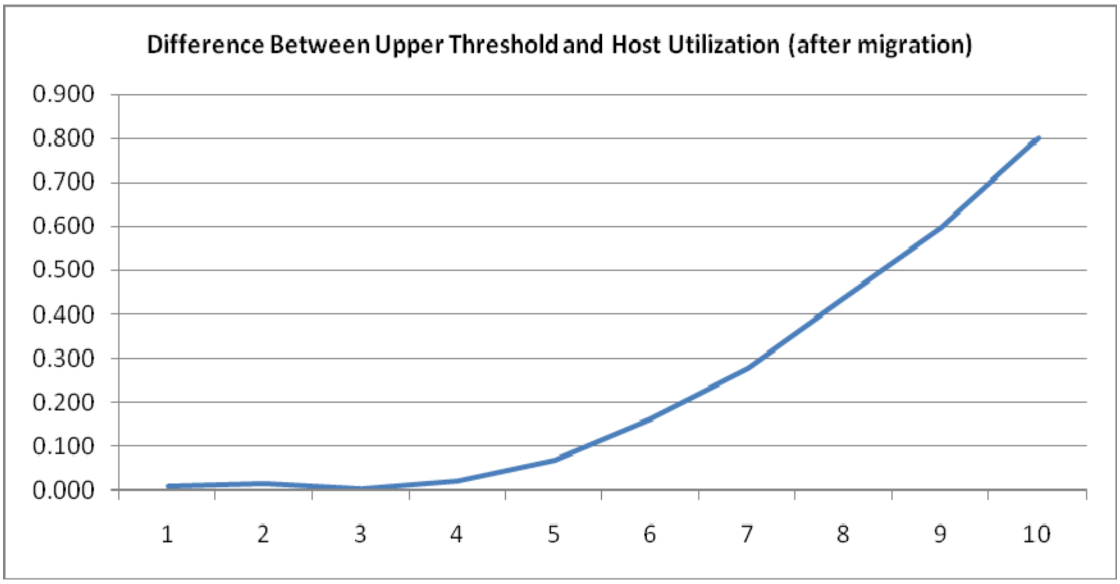

Figure 2. Difference between Upper Threshold and Host Utilization (after migration)

In our process of measuring migration time (*MinMigrTime*) mentioned in abstract, we carried out experimentation to calculate migration time for all possible options of number of VMs to be migrated. The results for various options of VMs are shown in Table 4 and the same has been depicted in figure 3.

TABLE 4. NUMBER OF VMS TO BE MIGRATED (CONSIDERING MIGRATION TIME)

Number of VMs to Migrate	**Actual VMS to Migrate (VM Ids)**	**Migration Time (µS)**
1	**210**	**0.0016**
2	204,207	0.0018
3	204,207,209	0.0024
4	203,204,205,209	0.0036
5	203,204,205,207,209	0.0048
6	203,204,205,206,207,209	0.0064
7	203,204,205,206,207,209,210	0.008
8	201,203,204,205,206,207,209,210	0.0098
9	201,203,204,205,206,207,208,209,210	0.0116
10	201,202,203,204,205,206,207,208,209,210	0.0141

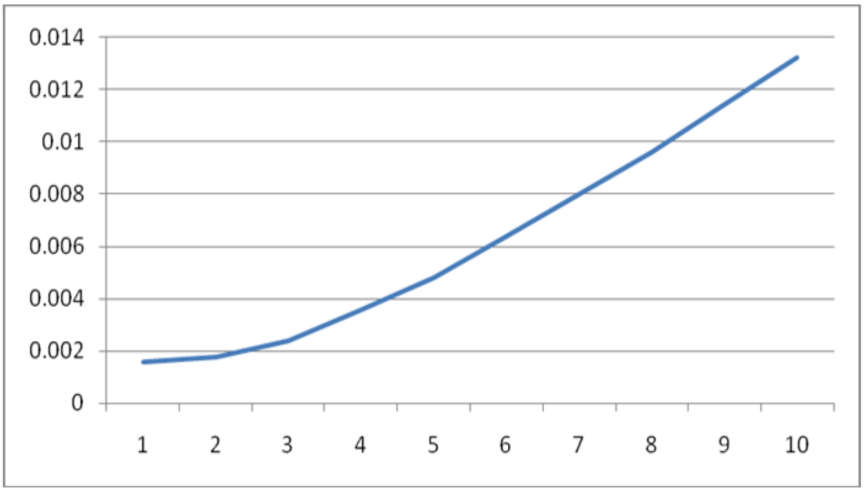

Figure 3. Migration Time (µs)

As show in figure, the solution with number of VM to be migrated = 1 with VM ID: 210 and migration time 0.0016 µs (*MinMigrTime*), is the optimal one.

Hence, if we summarize our entire criterion mentioned in abstract, possible option selected in each of it, are depicted in Table 5. As can be seen from Table 5 that Solution 5 and Solution 6 are same, hence ignoring one of them would left us with total 5 solutions.

TABLE 5. SUMMARIZED OPTIMAL SOLUTION FOR VM MIGRATION FOR HOST ID 2

No	**Criteria**	**Optimal Solutions {VM}**	**MaxUtil**	**MinTotalMigr**	**MinMigrTime (µS)**
1	Maximum Host Utilization (*MaxUtil*)	Solution 1: {203,204,205}	**0.0016**	3	0.003
2	Number of VMs to be migrated (*MinTotalMigr*)	Solution 2: {201}	0.0567	**1**	0.0018
		Solution 3: {202}	0.0083	**1**	0.0025
		Solution 4: {208}	0.0970	**1**	0.0018
		Solution 5: {210}	0.0567	**1**	0.0016
3	Minimum Migration Time (MinMigrTime)	Solution 6: {210}	0.0567	1	**0.0016**

Using "(2)", with various combinations of α, β and γ, we try to generate the possible outcome for making the decision of selecting the optimal solution. We conducted the experimentation for various combinations mentioned as per Table 6.

TABLE 6. PAIRING OF OPTIMAL SOLUTIONS AND CONSTANT SELECTIONS

	α	β	γ	Selection Index (SI) (Using Eq. 2)				
				Solution 1	**Solution 2**	**Solution 3**	**Solution 4**	**Solution 5**
Case 1	0.5	0.3	0.2	0.492	0.488	0.691	0.280	0.501
Case 2	0.3	0.2	0.5	0.295	0.458	0.491	0.333	0.491
Case 3	0.2	0.5	0.3	0.197	0.536	0.566	0.453	0.556
Case 4	0.8	0.15	0.05	0.787	0.426	0.790	0.093	0.436
Case 5	0.15	0.05	0.8	0.148	0.616	0.679	0.553	0.619
Case 6	0.05	0.8	0.15	0.049	0.441	0.279	0.420	0.494

Figure 4 summarized the values of Table 6 in form of graph.

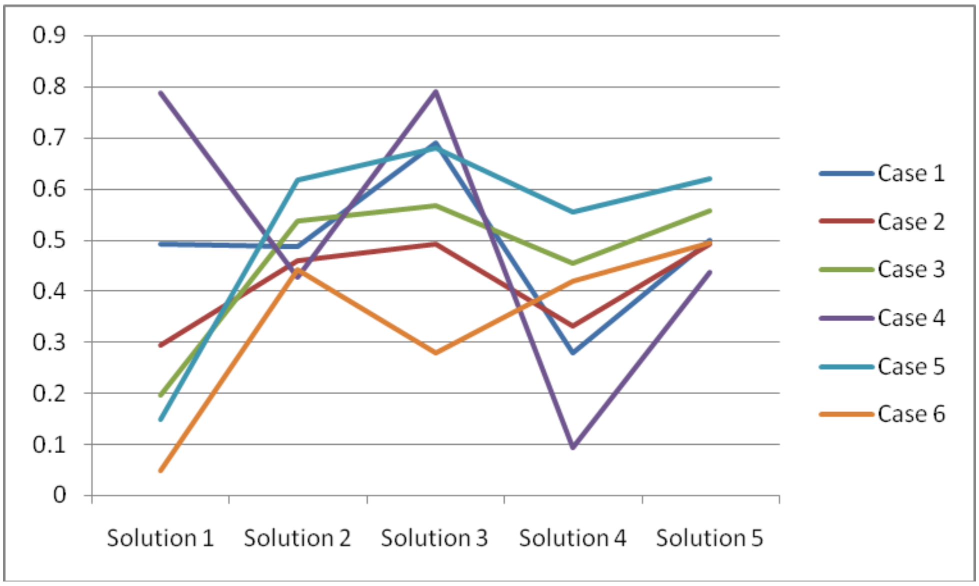

Figure 4: Pattern for Possible Options of Constants for Various Solutions Identified

It is evident from the figure 4 that for each solution identified (in Table 6) has different options for the cases selected. For instance, in case of Solution 1, Case 4 (i.e. α=0.8, β=0.15, γ=0.05) is an optimum one as it leads to higher Selection Index (SI). As mentioned earlier, HOST ID 4 is also overutilized. Table 7 summarizes the final optimal solution for it.

TABLE 7. SUMMARIZED OPTIMAL SOLUTION FOR VM MIGRATION FOR HOST ID 4

No	**Criteria**	**Optimal Solutions {VM}**	**MaxUtil**	**MinTotalMigr**	**MinMigrTime (µS)**
1	Maximum Host Utilization (*MaxUtil*)	Solution 1: {404,406,409}	<u>**0.00027**</u>	3	0.0028
2	Number of VMs to be migrated (*MinTotalMigr*)	Solution 2: {401}	0.00296	<u>**1**</u>	0.0018
		Solution 3: {402}	0.03790	<u>**1**</u>	0.0016
		Solution 4: {408}	0.00968	<u>**1**</u>	0.0018
3	Minimum Migration Time (MinMigrTime)	Solution 5: {402}	0.03790	1	<u>**0.0016**</u>

Similar results in form of table (such as 6) and figure (such as 4) can be generated for Host ID 4.

V. CONCLUSION

Increased usage of data centres in various sectors of ICT has led the researcher to consider the issue of energy consumed by them. Workload consolidation has been identified as one of the key directions to address the issue of energy consumption by optimally utilizing the data centre resources among the available applications while maintaining the users' requirement. Load balancing through VM migration is part of workload consolidation wherein selection of VM to be migrated is an open issue. In this research, we have addressed the issue of VM selection for migration from over-utilized servers. Multi-criterion policy has been defined and various experimentations have been conducted to understand the significance of various factors affecting the VM selection process. Further, defining the significance of individual factor is left to the user. In future, more criterions can be added to this process such as correlation among the VMs, priority of task etc. Further, this research has been implemented and tested on sample dataset which can be extended using real-world workload data provided as a part of the CoMon project, a monitoring infrastructure for PlanetLab.

VI. REFERENCES

[1] L. A. Barroso and U. Holzle, "The case for energy-proportional computing" IEEE Computer, vol. 40, no. 12, pp. 33– 37, December 2007.

[2] White Paper: Raffaele Giordanelli, Carlo Mastroianni, Michela Meo, Giuseppe Papuzzo, Andrea Roscetti, "Saving energy in data centers Through Workload Consolidation", 2007.

[3] A. Beloglazov and R. Buyya, "Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers", In Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science (Vol. 4). ACM, November 2010.

[4] A. Beloglazov and R. Buyya "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers", Concurrency and Computation: Practice and Experience, 24(13), 1397-1420, 2012.

[5] A. Horri, M.S. Mozafari and G. Dastghaibyfard, "Novel resource allocation algorithms to performance and energy efficiency in cloud computing", The Journal of Supercomputing, 69(3), 1445-1461, 2014.

[6] B. Wadhwa and A. Verma, "Energy and carbon efficient VM placement and migration technique for green cloud datacenters", in Contemporary Computing (IC3), 2014 Seventh International Conference on (pp. 189-193). IEEE, August 2014.

[7] J. Huang, K. Wu and M. Moh, "Dynamic Virtual Machine migration algorithms using enhanced energy consumption model for green cloud data centers",in High Performance Computing & Simulation (HPCS), 2014 International Conference on (pp. 902-910). IEEE, July 2014.

[8] Vikas Shukla and Guillaume Auriol, "Methodology for Determining Stakeholders' Criteria Weights in Systems Engineering" in Proceedings of the Posters Workshop at CSD&M 2013.

AUTHORS PROFILE

Nimisha Patel is currently working as an Associate Professor in Computer Engineering Department of Sankalchand Patel College of Engineering, Visnagar-Gujarat. She completed her Post Graduation (MTech-Computer Engineering) from National Institute of Technology (NIT), Surat-India and completed her under graduation (BE-Information Technology) from H.NG. University, Patan-India. Having more than 12 years of teaching experience, Ms. Patel has many research papers in various conferences & journals of international/national repute. Currently, Ms. Patel is pursuing PhD in the domain of Cloud Computing with resource allocation as specialization from Rai University, Ahmedabad. Few of her main subjects of interest are Cloud Computing, Data Structure, Distributed Computing, Parallel Processing, Information Security and Computer Networking.

Dr. Hiren Patel is currently working as a Professor and Head in Computer Engineering Department of LDRP Institute of Technology & Research, Gandhinagar-Gujarat. He completed his Ph.D. from National Institute of Technology (NIT), Surat with Cloud computing as the domain of research. He completed his post graduation from M. S. University, Baroda and graduation from S. P. University, V.V.Nagar. Having more than 14 years of teaching experience, Dr. Patel has many research papers in various conferences & journals of international repute. Few of his main subjects of interest are Computer Programming, Cloud computing, Parallel Processing, Computer Networking and Information Security.