

A Novel Approach for Submission of Tasks to a Data Center in a Virtualized Cloud Computing Environment

B.Santhosh Kumar

Dept. of Computer Science and Engineering
G. Pulla Reddy Engineering College
Kurnool, India

Dr. Latha Parthiban

Dept. of Computer Science
Community College, Pondicherry University
Pondicherry, India

Abstract—The submission of tasks to a data center plays a crucial role to achieve the services like scheduling, processing in a cloud computing environment. Energy consumption of a data center must be considered for task processing as it results in high operational expenditures and bad environmental impact. Unfortunately, none of the current research works focus on energy factor while submitting tasks to a cloud. In this paper a framework is proposed to select a data center with minimum energy consumption. The service provider has to register all the data centers in a registry. The energy consumed by task processing using virtualization and energy of IT equipments like routers, switches is calculated. The data center selection framework finally selects the data center with minimum energy consumption for task processing. The experimental results indicate that the proposed idea results in a less energy when compared to the existing algorithms for selection of data centers.

Keywords—Energy consumption; Virtualization; Data Center Selection Framework

I. INTRODUCTION

Cloud computing is growing predominantly because of the pay-as-you go model [1]. According to Gartner Report the cloud services expanded in to a \$210 billion market by the year 2016 [2]. Infrastructure maintenance is done by service providers while the actual development can be concentrated by the companies [3]. Data centers consume a lot of energy for storage and processing of data. The energy consumption of cloud infrastructure has been growing extensively and by the year 2020 it is expected to reach 1,963.74 terawatt-hours (TWh) [3]. The data centers growth rate is around 9 percent every year and hence the energy demands are also doubled in the last five years [4]. The heavy energy consumption of data Centers result in growth of operational expenses and also result in environment pollution due to excess of carbon emissions.

The tasks submission can be done in such a way that the data center with minimum energy consumption can be preferred over others for processing. To the best of our knowledge the existing algorithms don't concentrate on energy consumption. The algorithms often used in literature are outlined below:

1) **Spot electricity prices:** In this procedure the current electricity consumed by the data centers are taken in to

account and the tasks submission is done by predicting the future variations in the electricity prices. The data center with minimum electricity consumption is preferred. The drawback of this approach is it considers various non processing elements like lights, air conditioning units .., for selection of data centers.

2) **Shortest Distance First:** This algorithm forwards the request to the closest data center from the point of origin. The disadvantage is it does not take in to account the energy consumption which is relatively expensive when compared to the network elements like routers, bridges.., required to expedite the process of forwarding requests to the data centers.

3) **Round Robin Technique:** This algorithm chooses a data center randomly within a region where the request is originated. It might lead to a delay factor and the energy factor is also not considered.

To overcome the shortcomings in the existing approaches a data center selection framework (DCSF) is proposed which will calculate the energy used for processing of tasks. The execution time of the tasks and the energy consumed by the virtual machines are considered for calculation. The tasks are forwarded to the data center with minimum energy consumption. The procedure is simulated and compared with the existing algorithms. The results indicate that the proposed approach consumes less energy when compared to the others.

II. RELATED WORK

Many algorithms have been proposed for data center selection for task processing. Service broker routing policy is proposed in [5], [6] where the cloud analyst tool has been used. A probability based approach has been suggested in [7]. The research works in [8],[9] mainly focused on cost effective data center selection which take in to account the equipments needed to setup the data centers. A matrix based approach including the resources needed for data centers is discussed in [10]. The work in [11] signifies the expenditure spent on energy consumption for selection of data centers. A general data center selection approach based on factors like location, proximity is implemented in [12]. The algorithms in [13], [14] mainly concentrates on round robin and shortest job first approaches to select the data centers for processing. The virtualization technology which improves the task processing

capability is not given much importance in all these proposed works.

The importance of virtualization in cloud environment has been highlighted in [15]. Effective resource utilization using virtualization is discussed in [16], [17], [18] which focused on optimal allocation of resources to virtual machines. The scheduling of virtual machines for efficient energy consumption is formulated in [19] but it does not take into account the current energy being consumed by virtual machines. All the existing works concentrated on data center selection based on either random allocation or proximity and the energy consumption was given least importance. In the current work a framework is developed which focuses on the virtual machines consuming energy for processing of tasks.

III. PROPOSED WORK

In General the tasks generated by the users are transferred to the web servers situated at different locations which in turn route them to the data centers for processing. The current work forwards the tasks to a data center selection framework (DCSF) which calculates the energy consumed by different data centers and decides the one for which the tasks must be submitted for processing. The general depiction of the scenario is illustrated using Fig. 1.

TABLE I. DATA CENTER REGISTRY CONTENTS

Data Center ID (DCID)	No. Of Servers (NS)	No. Of Routers (NR)	No. Of Switches (NSW)	Total Energy Consumed (E _{Tot})

A user base (UB) is a region situated around the globe where a lot of users are submitting the tasks to the service provider.

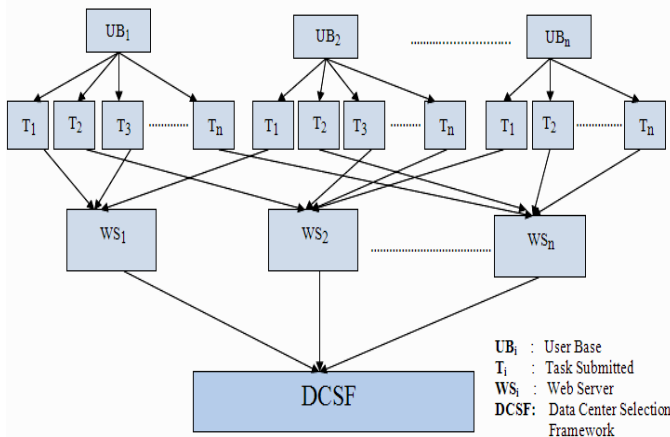


Fig. 1. Tasks Submission Process

Each data center is registered by the service provider in data center registry (DCR) as illustrated in Table I. A Data Center Id (DCID) is stored as a 16 bit number in which the first 4 bits represents service provider Id and the remaining 12 bits are allotted to data center id which should be a unique

number generated by the service provider. The information regarding number of servers, routers, switches..., in a data center has to be updated by the service provider. The total energy consumption is calculated and corresponding value is stored in DCR for further processing.

A. Data Center Selection Framework

The selection of data center is done by the Data Centre Selection Framework (DCSF) which is depicted in Fig. 2. Initially the tasks submitted to DCSF are stored in a Wait Queue. As soon as the tasks are removed from Ready Queue for processing then the tasks are allotted from Wait Queue into Ready Queue. A threshold value (Th) is considered depending upon the traffic received by the DCSF.

If the no. of tasks (λ) in Ready Queue > Th or the Time > 60s whichever condition is met earlier then the tasks are assigned to Task Allocator (TA). Then TA sends a request to Energy calculator Module (EC) where the total energy consumed by the data centers is calculated and the data center with minimum energy consumption is given the tasks for processing. The required information such as data center Id, no. of servers... is obtained from DCR. The server energy module (SEM) calculates the total energy consumed by the servers of data centers while the energy consumed by IT equipment is calculated by IT energy consumption (ITE) module.

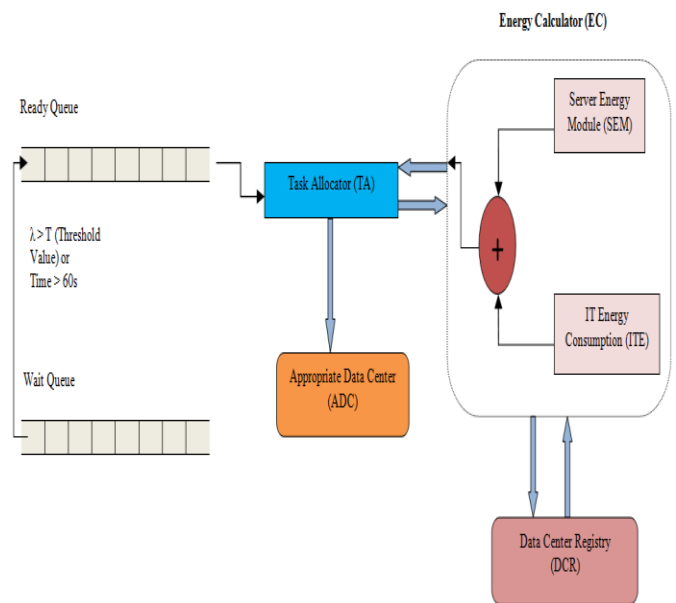


Fig. 2. Data Center Selection Framework

Finally the data center with minimum energy consumption is selected for task processing and the tasks are forwarded by TA to the appropriate Data center (ADC) depending upon its DCID.

a) Server Energy Module

In order to process the tasks efficiently and rapidly virtualization technique is employed in the servers of data centers. It also reduces the energy consumption of servers and increases the resource utilization. Multiple virtual machines are created on the same physical host thus increasing the

processing power of a server. Hence the energy consumption of the servers is obtained by the total energy consumed by execution of the tasks on virtual machines.

Let $S = \{S_1, S_2, S_3 \dots S_n\}$ be the set of servers located in a data centre DC_i . For each server S_k Let $VM_k = \{VM_{1k}, VM_{2k} \dots VM_{nk}\}$ be the set of virtual machines created. A decision variable x_{ijk} is set to 1 if task T_i is allotted to virtual machine VM_{jk} for processing. Let EC_{ijk} be the energy consumed by task T_i running on VM VM_{jk} and let ET_{ijk} be the execution time. The energy consumption rate of the VM is denoted by ECR_{ijk} and energy consumption EC_{ijk} can be calculated as follows:

$$EC_{ijk} = ECR_{ijk} \times ET_{ijk} \quad (1)$$

Where $ET_{ijk} = ft_{ijk} - st_{ijk}$ (finish time – start time)

Hence using (1) the total energy consumption can be calculated as follows:

$$TE_1 = \sum_{k=1}^{|S|} \sum_{j=1}^{|VM_k|} \sum_{i=1}^{|T|} x_{ijk} \times EC_{ijk} \quad (2)$$

Where $|T|$ Indicates Total no. of tasks submitted to VM_k

b) IT Energy Consumption Module

The no. of switches (NSW) and routers (NR) is obtained from DCR and the energy consumption is obtained by utility meter. The total energy consumed by IT equipments is given by TE_2 by equation (3)

$$TE_2 = NSW \times E_{SW} + NRT \times E_{RT} \quad (3)$$

Where E_{SW} is the avg. energy consumed by a switch and E_{RT} is the avg. energy consumed by a router of the data center. The output obtained by the two modules is combined to get the total energy (E_{Tot}) of the data center.

$$E_{Tot} = TE_1 + TE_2 \quad (4)$$

The total energy consumed (E_{Tot}) is updated in the DCR and the data center with the minimum E_{Tot} is selected as the appropriate data center (ADC) for tasks submission. The entire procedure can be illustrated in the following algorithm.

B. Data Center Selection Algorithm

```

START
  For Each Service Provider SP
    Register Each Data Centre  $DC_i$  in Data Centre Registry
    Registry DCR as in Table 1
  End For
  Initialize Ready Queue  $RQ \leftarrow Null$ ,
  Size  $(RQ) \leftarrow$  Threshold Value  $(Th)$ ,
  Wait Queue  $WQ \leftarrow Null$ 
  For Each Task  $T_i$  submitted by User Base  $UB_i$ 
    Assign  $WQ \leftarrow T_i$ 

```

If RQ is Null **Then** Assign $RQ \leftarrow WQ$

Until no. of tasks $(\lambda) >$ Size (RQ)

End If

End For

If $\lambda >$ Size (RQ) or Time $>$ 60s

Request Task Allocator TA to find Data Centre (ADC). TA assigns the request to Energy Calculator (EC) which retrieves each T_i from R and Perform the following computations:

$$|S| \quad |VM_k| \quad |T|$$

$$TE_1 = \sum_{k=1}^{|S|} \sum_{j=1}^{|VM_k|} \sum_{i=1}^{|T|} x_{ijk} \times ECR_{ijk} \times ET_{ijk}$$

$$TE_2 = NSW \times E_{SW} + NRT \times E_{RT}$$

$$E_{Tot} = TE_1 + TE_2$$

$$ADC \leftarrow DCID (\min (E_{Tot}))$$

End If

For Each Task T_i , TA does

Assign $ADC \leftarrow T_i$ for Processing

End For

END

IV. PERFORMANCE ANALYSIS

A. Analysis using real world trace logs

The proposed algorithm is evaluated by using the real world trace obtained from Google cloud trace logs [20]. The log contains information regarding 25 million tasks that were submitted for a period of 29 days. The analysis is done only by considering the tasks submitted to the cloud system in the first 5 hours of the day number 18. Fig. 3 illustrates the count of tasks submitted in every 60 seconds in those 5 hours. The count of the tasks is over 200 thousand.

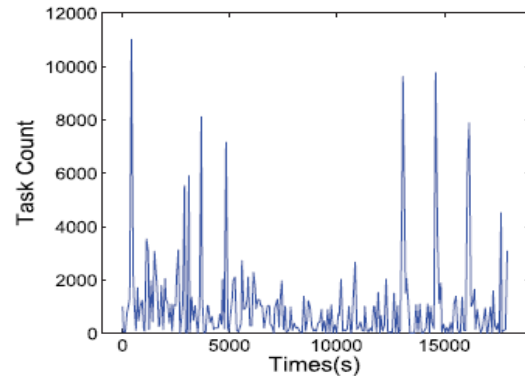


Fig. 3. Tasks Submitted to Cloud

The execution times of the tasks have been calculated by obtaining the start time and finish time from the logs. The sample data for our analysis is depicted in Table II. In order to understand easily the data center Ids are assumed as $DC1 \dots DC5$, instead of binary format. The units of energy consumption are in KWh. The metrics considered for analysis

of the algorithms are Total Energy Consumption (TOT) and Energy Consumption per Task (ECT).

TABLE II. DATA FOR SIMULATION

Data Center ID (DCID)	No. Of Servers (NS)	No. Of Routers (NR)	Energy Consumed per Router (E_{Rr})	No. Of Switches (NSW)	Energy Consumed per Switch (E_{Sw})
DC1	45	35	1.5	300	2.5
DC2	32	23	1.75	225	2.2
DC3	48	28	1.9	143	2.7
DC4	38	32	2.1	345	2.1
DC5	40	27	2.3	173	2.9

The proposed work Data center selection framework (DCSF) is compared with the Shortest Distance First (SDF) and Round Robin (RR) algorithms and the results are depicted in Fig. 4 and Fig. 5. When compared with the existing algorithms the total energy consumption is far less in the proposed approach. The data center DC4 is assumed to be shortest distance data center when compared to the others.

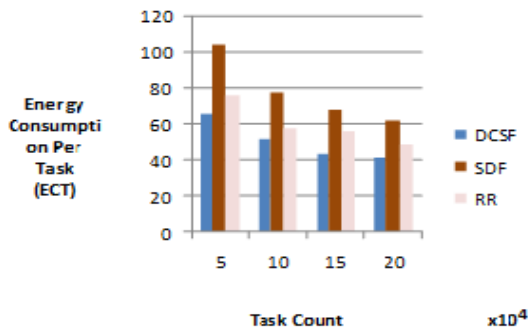


Fig. 4. Analysis using ECT metric

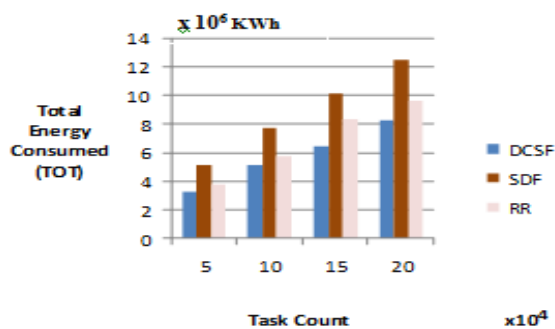


Fig. 5. Analysis using TOT metric

B. Analysis Using Node Red

The requests are generated as shown in Fig. 3 using Apache JMeter tool. All the threads generated by JMeter are created using algorithms of computer science like graph theory, data structures..., and submitted to three data centers.

```

[{"id": "5dd2be7.7a22d4", "type": "serial-port", "z": "", "serialport": "/dev
/ttyUSB0", "serialbaud": "57600", "databits": "8", "parity": "none", "stopbits": "1", "newline": ""
", "bin": "false", "out": "char", "addchar": "false", {"id": "5d9f1647.8da548", "type": "serial
in", "z": "d1767f1.cc1968", "name": "CurrentCost
Envir", "serial": "5dd2be7.7a22d4", "x": 298, "y": 161, "wires": [{"2e38cccb.02c6b4"}]},
{"id": "1656c54c.43f1eb", "type": "function", "z": "d1767f1.cc1968", "name": "extract
watts", "func": "/* sample message\n(Object) {\n \"msg\": {\n \"src\": [\"CC128-
v0.11\"]\n, \"deb\": [\"00600\"]\n, \"time\": [\"19:04:49\"]\n, \"tmpr\": [\"18.3
\"]\n, \"sensor\": [\"0\"]\n, \"id\": [\"02902\"]\n, \"type\": [\"1\"]\n, \"ch1\":
[\"watts\": [\"00023\"]]\n}\n}\n*/\n\nvar reading = msg.payload\n\n//ignore if history
packet\n\nif (reading.msg.hist) (return null;)\n\nmsg.payload = {};\n\nmsg.payload.sensor =
reading.msg.sensor[0]*1;\n\n// This is the temperature of the Envir base station, not the
sensor.\n\nmsg.payload.temperature = reading.msg.tmpr[0]*1;\n\nmsg.payload.watts =
reading.msg.ch1[0].watts[0]*1;\n\nreturn msg;
", "outputs": 1, "noerr": 0, "x": 625, "y": 161, "wires": [{"d8e625b9.46a778"}]},
{"id": "d8e625b9.46a778", "type": "debug", "z": "d1767f1.cc1968", "name": "", "active": false, "compl
ete": false, "x": 805.5, "y": 161, "wires":
[[]], {"id": "2e38cccb.02c6b4", "type": "xml", "z": "d1767f1.cc1968", "name": "", "x": 466.5, "y": 161, "
wires": [{"1656c54c.43f1eb"}]}

```

Fig. 6. Node Red flow to read energy from a Server

The data centers are implemented using three Dell Power Edge C5220 Micro servers represented as DC1, DC2, and DC3. The data center capabilities such as virtualization, storage management are implemented using Windows Server 2012 R2 data center Operating System which is installed in the data centers. The Raspberry Pi chips are incorporated in each of the servers in the data centers so as to read the energy consumption.

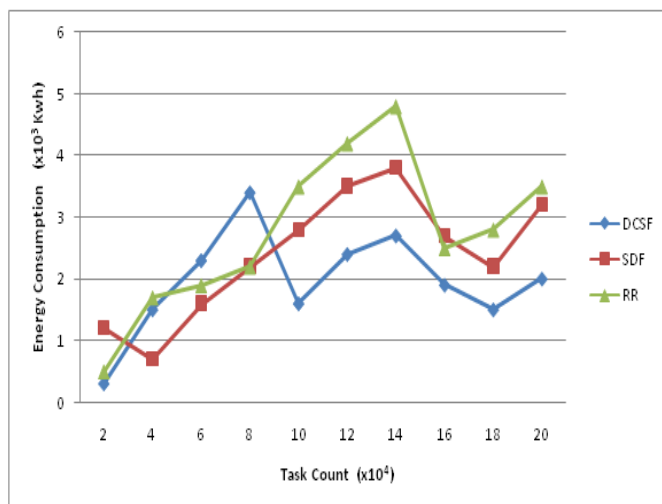


Fig. 7. Analysis using Node Red

The results obtained are illustrated in Fig. 7 which shows that the total energy consumed by the proposed framework is less than the energy consumption of the remaining algorithms. While the remaining algorithms like Shortest Distance First (SDF), Round Robin (RR) consume less energy at some instances, on an average the proposed framework DCSF results in low energy consumption.

V. CONCLUSION

In this paper a framework is implemented for selection of data centers based on minimum energy consumption of data centers for task processing. The total energy is calculated by summation of the server energy consumption and IT equipment energy calculation. The server energy is obtained by combining the energy consumed by virtual machines for task processing. The results are obtained by simulation of real

world cloud trace logs and Node-red tool. The obtained results indicate that the proposed framework outperforms the already existing algorithms for selection of data centers. This work can be further extended by implementing the concept of Internet of Things using the IBM Bluemix tool.

REFERENCES

- [1] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg and I. Brandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the Fifth Utility," *Future Generation Computer Systems*, Vol.25, 2009, pp. 599-616.
- [2] "Gartner. Forecast overview: Public cloud services," Worldwide 2013-2019, 1Q15 update.
- [3] Greenpeace international, "Make IT Green, cloud computing and its contribution to climate change," 2010.
- [4] IBM, "IBM and Cisco: Together for a World Class Data Centre," 2013.
- [5] A.M. Manasrah, T.Smadi and A.ALmomanA, "A variable service broker routing policy for data centre selection in Cloud Analyst," *Journal of king saud university*, Vol.28, 2016, pp.1-13
- [6] D.Kapgate, "Efficient Service Broker Algorithm for Data Centre Selection in Cloud Computing," *International Journal of Computer Science and Mobile Computing*, Vol.3, 2014, pp.355-365.
- [7] S.Yan, X.Wang, M.Razo, M.Tacca and A.Fumagalli, "Data centre selection: A probability based approach," in proceedings of 16th International Conference on Transparent Optical Networks, 2014, pp.1-5.
- [8] H.Xu and B. Li, "Cost Efficient Data Centre Selection for Cloud Services", in proceedings of 1st IEEE International Conference on Communications in China, 2012, pp.51-56.
- [9] P.M.Rekha and M.Dakshayini, "Cost Based Data Centre Selection Policy for Large Scale Networks," in proceedings of International Conference on Computation of Power, Energy, Information and Communication, 2014, pp.18-23.
- [10] A.Jaikar, G.Ryoon Kim and S.Young Noh, "Matrix-based Data Centre Selection Algorithm for a Federated Cloud," *International Journal of Multimedia and Ubiquitous Engineering*, Vol.9, 2014, pp.143-148.
- [11] K.S.Swe Myint and T.Thein, "Energy and Cost Efficient Data Centre Selection for Cloud Services," in proceedings of 3rd International Conference on Computational Techniques and Artificial Intelligence, 2014, pp.35-38.
- [12] H. Xu and B.Li, "A General and Practical Data Centre Selection Framework for Cloud Services," in proceedings of IEEE 5th International Conference on Cloud Computing, 2012, pp.9-16.
- [13] V.Sharma, Dr. Rakesh and S.Bol, "Round-Robin Data Centre Selection in Single Region for Service Proximity Service Broker in Cloud Analyst," *International Journal of Computers and Technology*, Vol.4, 2013, pp.254-260.
- [14] D.Chudasama, N.Trivedi and R.Sinha, "Cost effective selection of Data centre by Proximity-Based Routing Policy for Service Brokering in Cloud Environment," *International Journal of Computer Technology and Applications*, Vol.3, 2012, pp.2057-2059.
- [15] K.Shyamala and R.T.Sunitha, "An analysis on efficient resource allocation mechanisms in cloud computing," *Indian Journal of Science and Technology*. Vol.8, 2015, pp.814-821.
- [16] I. Goiri, J.L. Berral, J.O. Fit_o, F. Juli_a, R. Nou, J. Guitart, R. Gavald_a and J. Torres, "Energy-Efficient and Multifaceted Resource Management for Profit-Driven Virtualized Data Centres," *Future Generation Computer Systems*, Vol.8, 2012, pp.718-731.
- [17] Z.Xiao, W.Song and Q.Chen, "Dynamic Resource allocation using virtual machines for cloud computing environment," *IEEE Transactions on Parallel and Distributed Systems*, Vol.24, 2013, pp.1107-1117.
- [18] C.Papagianni, A.Leivadeas, S.Papavassiliou, V. Maglaris, C.Cervello and A.Monje, "On the optimal allocation of resources in cloud computing Networks," *IEEE Transactions on Computers*, Vol.62, 2013, pp.1060-1071.
- [19] D.Jiankang, W.Hongbo, L.Yangyang and C.Shiduan, "Virtual machine scheduling for improving energy efficiency in IaaS cloud", *China communications*, Vol.11, 2014, pp.1-12.
- [20] GoogleClusterDataV2[Internet], 2011 Available from: http://code.google.com/p/googleclusterdata/Wiki/ClusterData2011_1