RESEARCH ARTICLE OPEN ACCESS

Energy Reduction Techniques in Cloud Computing: A Review

Jasleen kaur, Navpreet Kaur Walia
Department of Computer Science and Engineering
Sri Guru Granth Sahib World University
Punjab - India

ABSTRACT

The Cloud computing is a new computing technology which aims at offering reliable, adaptive and Quality of Service (QoS) based computing environments for IT. One of the recent major concerns in cloud computing environment is inefficient and uneconomical usage of energy in data processing, storage and communication. This is harmful for the environment due to carbon emission. Therefore, green IT is required to save the environment. The green cloud computing approach is part of green IT which aims to reduce the carbon footprint of datacenters by reducing their energy consumption. This paper reviews various energy saving approaches for Data Centers in cloud computing and also discusses various approaches proposed in the previous research works in this field. The paper presents a study focusing on energy consumption by data center servers and networks and energy efficient approaches in cloud computing. Based on the existing studies, it is seen that the servers consumes the largest amount of energy in data centers.

Keywords:- Cloud, Energy, Green IT, Data centers.

I. INTRODUCTION

Cloud Computing is a technology that is used to provide resources over the internet. The resources can be any kind of application (like video conferencing applications, emails etc.), platforms (like java platform), or infrastructure facilities (like physical or virtual servers instances). The concept of cloud computing has been explained in the following figure (Fig 1).

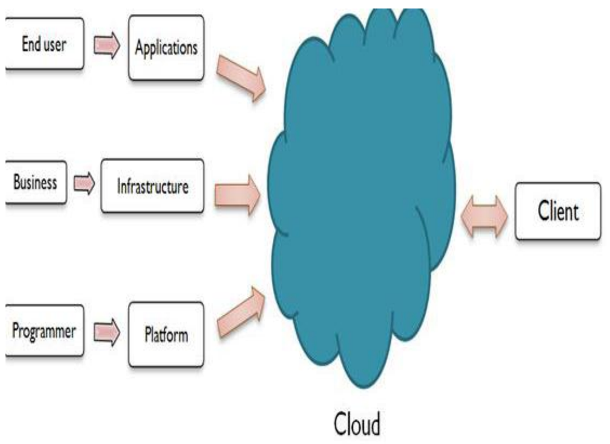

Fig 1. Concept of Cloud

National Institute of Standards and Technology [4] defines three service models for Clouds .i.e. SaaS, PaaS and IaaS. SaaS (or Software as a Service)model provide applications to the cloud users over the internet. PaaS (or Platform as a Service model) provides platform to developers to develop their applications. IaaS (or Infrastructure as a Service),alsoknown as Hardware as a Service provides infrastructure facilities to the customers. NIST also defines five essential characteristics and four deployment models of Cloud Computing [4].

The various features of cloud computing are On-demand service, Infinite network access, Location independence and resource pooling, Scalable resources, Service is measured.

And the three models of deployment are: public model, private model, community model and hybrid model. Thus establishing a Cloud requires installation of a large number of computing resources in a single data center.

II. OVERVIEW OF ENERGY REDUCTION TECHNIQUES

In general, green cloud computing can be implemented via three approaches: software optimization [2] [19] [11] [20] [21] , hardware optimization [23][24][25][26], or network optimization [6][27][28][29][30] in order to reduce the power consumption, as illustrated in Fig. 2.

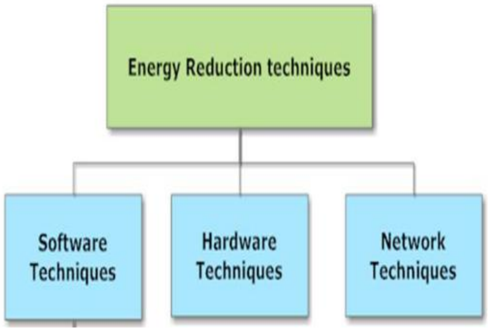

Fig. 2 Energy Reduction Techniques

III. SOFTWARE TECHNIQUES

There are many ways to achieve energy efficiency in clouds, some of them are stated below:

1. *Use the virtualization technology*: By using the Virtualization technology, we are able to create virtual instances of a single physical node. Each of these virtual instances is capable to act as a separate host for the cloud jobs. This helps to reduce the number of physical hosts required to execute the tasks as well as it helps to increase the utilization of the physical server.
2. *Switch the idle hosts to other low power consuming modes*: The hosts that are idle can be either switched off or they can be changed to some lower power consuming modes like sleep or hibernate. This helps to reduce a lot of power consumption.
3. *Allowing the live migration of VMs:* The Vms can be allowed to migrate from one virtual machine to the other. This can be done to make the utilization of one host to maximum level and to make an underutilized host completely idle so that we could turn it off or change it to some low power mode.

A. *Reducing Server Consumed Energy*

The energy consumption of servers can be decreased by reducing the number of active servers. This is usually implemented by scheduling optimization, which is a common approach for green clouds and is considered [31] more efficient than hardware optimization, in terms of cost, consumed hiaa hoia oz aooiorliltr. It depends on finding a suitable mapping between requests for VMs and physical servers to minimize the amount of consumed power [31].

One of the important issues for energy efficiency in virtualized cloud environments is where to place new VM requests within the physical servers. In [2], they proposed a heuristic-based energy-efficient approach for VM placement in cloud based datacenter which relies on statistical analysis of historical data. It uses the multiple correlation coefficient (MCC) method; i.e. measuring the strength of association between a given variable and other variables; to select the server that provides a suitable trade-off between power efficiency and SLA violation.

Higher the correlation coefficient of CPU utilization between selected VMs running on a server, the higher the risk of SLA violation in the datacenter. They used the CloudSim 3.0 toolkit to simulate a datacenter with heterogeneous physical hosts and computed the energy consumption. However, this algorithm requires information from the hardware level [2].

An inspired approach is proposed in [19] based on the behavior of real ants. It uses the Ant Colony Optimization (ACO) meta-heuristic for VM placement. It aims to minimize the number of active servers by maximizing the resources utilization. VM placement is computed dynamically according to the current server load. In this algorithm, each ant (server) receives all VMs, and starts scheduling them to the servers. After all ants have built their solutions, the solution with lowest value for objective function is chosen. The pheromone trails are updated for only this solution. They have developed their own java-based simulation toolkit [19]. . It also wastes a lot of resources as each ant must compute its own solution, but then only the one with the lowest value of objective function will be considered.

B. *Server Consolidation Algorithm*

A server consolidation algorithm (Sercon) is proposed in [11]. It aims to reduce the energy consumption in homogeneous datacenters by minimizing the number of active servers. The algorithm inherits some properties from First- and Best-Fit (FF and BF) bin packing problems. First, it sorts the servers in decreasing order based on their load. Then the VMs in the least loaded server become candidates for migration. Thereafter, those VMs are sorted in decreasing order based on their weights. After that, those VMs are allocated one-by-one to the most loaded servers. Therefore, the least loaded server will be idle. So the number of running servers is reduced by switching off idle servers, reducing energy consumption. A simulation software is developed using the .NET 3.5 framework to evaluate an experiment of the proposed method. However, the algorithm is fully centralized, considers homogenous servers, and prevents the server's processor from being fully utilized.

VM migration technique can be used to optimize scheduling. This focuses on transferring VMs between servers via the network. It is used as a solution for improving energy efficiency, by consolidating the VMs on fewer physical servers [20][21]. In [20] both scheduling and VM migration methods were used to reduce the energy consumed by servers. It proposed the Energy-aware Scheduling algorithm using Workload-aware Consolidation Technique (ESWCT).

- The algorithm aims to consolidate the VMs in minimum amount of servers based on balancing the integrated resources (processor, memory and network bandwidth) which are shared concurrently among users in cloud datacenters. It considers heterogeneous workloads of various resource consumption characteristics.
- The aim of this algorithm is to reduce the power consumption by improving resources utilization based on the fact that heterogeneous workloads have different resource consumption characteristics
- . The algorithm is centralized and consists of two phases: the first phase shows where to place the VM to get a better balanced utilization of resource among physical servers.

In [21], they propose an approach for VM placement and migration to deal with both over-utilized and under-utilized servers. In the VM placement, they applied a modification of the Best Fit Decreasing algorithm called Power Aware Best Fit Decreasing (PABFD), to allocate a new request for a VM to a server that provides the least increase in power consumption caused by the allocation.

- The PABFD method is applied by sorting all VMs in decreasing order of their current processor utilizations and allocates each VM to the server that provides the least increase of power consumption caused by the allocation. This migration approach is proposed to overcome the problem of under loaded servers. The system finds the server with the minimum utilization compared to the other servers, all of its VMs are selected for migration.
- The target server is selected based on a Best Fit Decreasing heuristic: the migrating VMs are sorted in decreasing order by processor utilization and placed in the server that provides the least increase in power consumption, the source server is then switched to the sleep mode (power-saving mode) once all migrations have been completed. If all the VMs from the source server cannot be placed on other servers, the server is kept active.

The CloudSim toolkit simulation tool to implement the proposed algorithm. In order to evaluate the efficiency of the algorithms, the total energy consumption and the number of VM migrations are used as performance metrics.

IV. HARDWARE TECHNIQUES

Other technique reduce the consumed energy by utilizing flexible hardware that varies the server computing capability via controlling the frequencies and voltages in the server, which affects the energy consumption [31]. However, as with all other hardware techniques, this approach to green cloud is costly and suffers from poor scalability because of the special hardware requirements. The various hardware approaches are:

- *Power-aware scheduling algorithm(dvfs)* :A power-aware scheduling algorithm is presented in [23]. It implements Dynamic Voltage Frequency Scaling (DVFS) technique, which is applied with a number of special processors that can to operate at different voltage and frequency levels. It selects the appropriate supply voltages and frequencies of processing elements to minimize energy consumption without violating the SLA, based on the VMs workload. Each VM is allocated to the First Fit server, and each server applies the DVFS to save the energy while complying with the SLA requirements. The result shows a reduction in energy consumption without violating the SLA, and is compared with a non-power aware algorithm. It is implemented by using CloudSim toolkit and it is provided as an example in the simulator.
- *Task level DVS approach:* The study in [24] follows the same approach but they consider the SLA based on the task level (task deadline), and scaling only the supply voltage. The scheduling algorithm applies Dynamic Voltage Scaling (DVS) to save energy while testing the ability of each scheduled task to meet its deadlines. They consider two DVS scheduling policies: one a space-shared policy and the other a time-shared policy. They simulate the proposed algorithms by using GridSim toolkit. The DVS-server is required to control its supply voltages, so it needs special or additional hardware.
- *Energy-Aware Resource Efficient workflow Scheduling under Deadline constraint (EARES-D):*The study in [25] applied DVFS technique to find the optimal frequency for each task of a scientific workflow without affecting its performance. A multi-step heuristic workflow scheduling algorithm is proposed, namely Energy-Aware Resource Efficient workflow Scheduling under Deadline constraint (EARES-D). In the first phase, they calculate the estimated earliest completion time for a workflow, in all datacenters. Then the optimal frequency for executing each task is determined by scaling down the processor frequency under the deadline constraint. The datacenter is selected based on the first and second phase. Thereafter, the task is forwarded to the selected datacenter for scheduling. The resource utilization rate is improved by reusing

VM and shrinking the idle time between tasks if the deadline is still guaranteed. They used CloudSim toolkit as a simulator tool to evaluate the scheduling algorithm. The simulation results showed an improvement in energy consumption and resources utilization rate.

V. NETWORK TECHNIQUES

The communications between VMs consumes energy in the datacenter [6]. Reducing the network traffic between servers reduces energy consumption. The studies [6][27][28][29][30] consider the network traffic of the VMs placements to reduce the energy consumption.

In [7], two heuristics for VMs migration are presented based on the communication graph and other resource requirements such as processor, memory etcetera. The communication graph is represented as a weighted graph. The weight for each edge in the communication graph shows the amount of traffic between two VMs. So the connected component means those VMs communicate with each other while disconnected components means there is no network traffic between these components. The algorithm identifies the under-loaded servers and the heavily-loaded servers. Then it identifies the physical servers with sufficient residual capacity and sorts them in ascending order according to load. From the lightly loaded servers, it identifies the set of VMs whose load can be accommodated by these physical machines and constructs the communication graph of those VMs. After that, it sorts the components in decreasing order of their size. The algorithm migrates the largest and least connected component first. Each component is migrated as a whole to a single physical machine based on the load of the VMs and the residual capacity of the target physical machine. So the VMs with high communications with each other will be in one server. If this is not possible, either a modified breadth-first search algorithm or a modified Prim's maximum spanning tree algorithm is used to partition the VMs. Then the partitions are migrated to physical machines in proximity to each other based on their distance matrix. Thus, the network traffic between the servers is reduced resulting in less power consumption in the datacenter. They didn't implement these algorithms for the evaluation.

- Study [27] optimizes the VM placements by consolidating the VMs to the minimum number of servers, and reducing the network traffic between those, to decrease the energy consumption. It consolidates the VMs with high communication flow together in order to reduce the network traffic between racks, number of active servers and number of active network elements (links and switches). It evaluated in terms of simulation to estimate the number of active servers and intra-rack traffic.
- Datacenter Energy-efficient Network-aware Scheduling (DENS) is proposed in [28]. It aims to reduce the energy consumption in a datacenter by optimizing the tradeoff between task consolidation and traffic pattern distribution. The proposed DENS selects the best-fit server to execute a job based on weighted computational function that considers the load and the communicational potential at server, rack, and module levels. The proposed function converges VMs towards the maximum loaded server in the least-utilized rack with low network traffic.
- The study in [30] proposes VM placement algorithm that aims to provide a balance between server energy consumption and network energy consumption. It considers the datacenter as a dependency graph, similar to [6]. It employs fuzzy logic to combining those two conflicted objectives.

In summary, network optimization techniques provide a reduction in the energy consumption with the ability to meet the SLA. On the other hand, a datacenter is usually constructed with a fixed network topology, which limits the scalability and the flexibility in the datacenter. This approach needs to be aware of network topology to decide the flow route as in [28].

VI. CONCLUSION

In general, the growth of cloud computing has led to uneconomical energy consumption in data processing, storage and communication. The massive energy consumption is unfriendly to the environment because of the huge carbon footprints of the datacenters. Therefore, green cloud computing is required to support the environment. Green computing produces environmental-friendly and cost-efficient cloud computing by using computing resources more efficiently.

This paper overviews the GCC approaches and classifies them. This classification assists in comparisons between GCC approaches by recognizing the key implementation techniques and the related issues Three approaches can be followed to implement the green cloud computing: software optimization, hardware optimization, and network optimization.

The software optimization is easy to implement and most scalable, usually not requires special network topology or special hardware. But in the software optimization techniques, SLA compliance and energy consumption have a negative correlation. Hardware optimization provides a reduction in the energy consumption while complying with the SLA. On the other hand, it more costly and has a limitation in scalability because of the special hardware requirements. Network optimization techniques can reduce energy consumption while complying with the SLA. But it needs to be aware of the network topology and can applied only in a specific network topology, which limits its scalability and the flexibility.

REFERENCES

[1] G. von Laszewski and L. Wang, "GreenIT service level agreements," in Grids and Service-Oriented Architectures for Service Level Agreements, P. Wieder, R. Yahyapour and W. Ziegler, Eds. New York, NY, USA: Springer, 2010, pp. 77–88.

[2] N. Kord and H. Haghighi, "An energy-efficient approach for virtual machine placement in cloud based data centers," in Proc. 5th Information and Knowledge Technology Conf., 2013, pp. 44–49.

[3] B. Hoare, Animal Migration: Remarkable Journeys in the Wild, Berkeley, CA, USA: Univ. of California Press, 2009.

[4] Peter Mell,Timothy Grance, "The NIST Definition of Cloud Computing (Draft)," Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology ,Gaithersburg, January 2011.

[5] A. Beloglazov, J. Abawajy and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," Future Generation Computer Systems, vol. 28, no. 5, pp. 755–768, 2012.

[6] G. S. Akula and A. Potluri, "Heuristics for migration with consolidation of ensembles of virtual machines," Proc. Communication Systems and Networks (COMSNETS), 2014 6th Int. Conf., pp. 1, 4, 6–10.

[7] S. F. Smith, "Is scheduling a solved problem?" in Multidisciplinary Scheduling: Theory and Applications, G. Kendall, E. K. Burke, S. Petrovic and M. Gendreau, Eds. Nottingham, UK: Springer, 2005, pp. 3–17.

[8] J. Whitney and P. Delforge, Data Center Efficiency Assessment. New York, NY, USA: Natural Resources Defense Council, 2014.

[9] "Data center locations." [Online]. Available: http://www.google.com/about/datacenters/inside/locations/index.html. [Dec. 25, 2015].

[10] J. Pearn. [Online]. Available: https://plus.google.com/+JamesPearn/posts/VaQu9sNxJuY. 2012. [Dec 25, 2015].

[11] A. Murtazaev and S. Oh, "Sercon: Server consolidation algorithm using live migration of virtual machines for green computing," IETE Technical Review, vol. 28, no. 3, pp. 212–231, 2011.

[12] S. Marston et al., "Cloud computing – The business perspective", Decision Support Systems, vol. 51, pp. 176–189, Apr., 2011.

[13] M. Rouse, "Data center definition." [Online]. http://searchdatacenter.techtarget.com/definition/data-center. Aug. 2010. [Mar. 8, 2015].

[14] J. Sahoo, S. Mohapatra and R. Lath, "Virtualization: A survey on concepts, taxonomy and associated security issues," Proc. Computer and Network Technology (ICCNT), 2010 2nd Int. Conf., Bangkok, 2010, pp. 222–226.

[15] M. Gawali, "ESDL – Virtual machine administration." [Online]. http://mahenswap.blogspot.com/2014/06/virtual-machine-administration.html. June 2014. [Mar. 12, 2015].

[16] S. Agarwal, A. Datta and A. Nath, "Impact of green computing in IT industry to make eco friendly environment," Journal of Global Research in Computer Science, vol. 5, no. 4, pp. 5–10, Apr. 2014.

[17] A. Elgelany, "Integrated framework for green ICT: Energy efficiency by using effective metric and efficient techniques for green data centres", Ph.D. dissertation, Comp. Sci., Sudan University of Science & Technology, 2015.

[18] "Secretary Chu announces $47 million to improve efficiency in information technology and communications sectors," [Press release]. Washington, DC, USA: US Dept. of Energy, Jan. 6, 2010. [Archived from the original on May 27, 2010].

[19] E. Feller, L. Rilling and C. Morin, "Energy-aware ant colony based workload placement in clouds," Proc. 2011 IEEE/ACM 12th Int. Conf. on Grid Computing, pp. 26–33.

[20] H. Li, J. Wang, J. Peng, J. Wang and T. Liu, "Energy-aware scheduling scheme using workload-aware consolidation technique in cloud data centres," Communications, China, vol. 10, no. 12, pp.114, 124, Dec. 2013. doi: 10.1109/CC.2013.6723884

[21] 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, vol. 24, no. 13, pp. 1397–1420, 2012.

[22] Anton Beloglazov_ and Rajkumar 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 Concurrency Computat.: Pract.Exper. 2012; 24:1397–1420

[23] R. N. Calheiros, R. Ranjan, A. Beloglazov1, C. A. F. De Rose and R. Buyya, "CloudSim: A toolkit for modeling and simulation of cloud computing environment and evaluation of resource provisioning algorithms," Software: Practice and Experience, vol. 41, no. 1, pp. 23– 50, 1ozs nrcc s oi: 10.1109/TC.2010.82

[24] K. H. Kim and J. Kim, "Power aware scheduling of bag-of-tasks applications with deadline constraints on DVS-enabled clusters," CCGRID, vol. 7, pp. 541–548, May 2007.

[25] F. Cao and M. M. Zhu, "Energy efficient workflow job scheduling for green cloud," Proc. 2013 IEEE 27th Int. Symp. on Parallel and Distributed Processing Workshops and PhD Forum, pp. 2218–2221.

[26] L. M. Zhang, K. Li and Y.-Q. Zhang, "Green task scheduling algorithms with speeds optimization on heterogeneous cloud servers," Proc. of the 2010 IEEE/ACM Int. Conf. on Green Computing and Communications & Int. Conf. on Cyber, Physical and Social Computing, Washington, DC, USA, pp. 76–80.

[27] T. Yapicioglu and S. Oktug, "A traffic-aware virtual machine placement method for cloud data centers," Proc. Utility and Cloud Computing (UCC) 2013 IEEE/ACM 6th Int. Conf., pp. 299–301.

[28] D. Kliazovich, P. Bouvry and S. U. Khan, "DENS: Data center energy-efficient network-aware scheduling," Proc. Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int. Conf. & Int. Conf. on Cyber, Physical and Social Computing (CPSCom), pp. 69–75. doi: 10.1109/GreenCom-CPSCom.2010.31

[29] W.-C. Lin, C.-H. Liao, K.-T. Kuo and C. H.-P. Wen, "Flow-and-VM migration for optimizing throughput and energy in SDN-based cloud datacenter," Proc. Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th Int. Conf., pp. 206–211. doi: 10.1109/CloudCom.2013.35

[30] D. Huang, D. Yang, H. Zhang and L. Wu, "Energy-aware virtual machine placement in data centers," Proc. Global Communications Conference (GLOBECOM), 2012, pp. 3243–3249.

[31] B. Gayathri, "Green cloud computing," Proc. Sustainable Energy and Intelligent Systems (SEISCON 2012), IET Chennai 3rd Int. Conf., pp. 1–5.

[32] H. A. Kurdi, "Personal mobile grids with a honeybee inspired resource scheduler," Ph.D. dissertation, ECE, Brunel University, London, UK, 2010.

[33] A. Beloglazov, R. Buyya, Y. C. Lee and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems," arXiv preprint: arXiv:1007.0066, 2010.