International Journal of Computer Science and Engineering Communications
Vol.3, Issue 3, 2015, Page.1141-1148
ISSN: 2347–8586
www.scientistlink.org

Energy-Aware Scheduling Using Workload Consolidation Techniques in Cloud Environment

[1]Sridharshini V, [2]V.M.Sivagami

[1] PG Scholar, [2] Associate Professor
Department of Information Technology
Sri Venkateshwara College of Engineering, Chennai, India.
[1]dharshini.it07@gmail.com, [2]vmsiva@svce.ac.in

Abstract: In cloud computing, a cloud is a managed pool of resources which provide on-demand services or computational resources to the remote users over a network. The resources are provided to users in the form of virtual machines and are possibly distributed and heterogeneous, running on the cloud environment over Internet. Energy-aware Scheduling algorithm and Energy-aware Live Migration algorithm reduces energy consumption in cloud environment. Both Algorithm use workload consolidation Techniques to make full use of the resources and investigate the problem of consolidating heterogeneous workloads and try to execute all Virtual Machines (VMs) with the minimum amount of Physical Machines (PMs), then power off unused physical servers to reduce power consumption. Energy-aware Scheduling Algorithm, Randomly turn on a physical server which is in sleep mode to active mode and place the VM to server. Energy-aware live Migration, VMs on underused physical servers to those which are mostly fully used that should be migrated to other unused physical server and is to optimize the current VM allocation and to migrate VMs. Multi-dimensional resources taken into consideration when placing VM and considering heterogeneous workloads. Both algorithms efficiently utilize the resources, and the multidimensional resources have good balanced utilizations, gives their promising energy saving capability.

***Keywords**:*
Energy-aware scheduling; Live Migration; Heterogeneous workloads; Workload-aware consolidation.

1.Introduction

Cloud computing is emerging with growing popularity and adoption. Cloud computing delivers an infrastructure, platform, and software as services to users in the form of leases. Consider an Infrastructure-as-a- Service (IaaS) cloud system. In this kind of cloud system, the resources are provided to users in the form of virtual machines running on the cloud provider's computing centre over Internet. And the end-users do not need to own any part of the infrastructure. Users only pay for the resources they have consumed, just as water or electricity. Various studies showed that cloud computing can reduce the infrastructure and IT management cost. Many cloud computing service providers including Amazon, Google, IBM and Microsoft are delivering cloud computing services. And many companies use cloud computing to provide more efficient services to their customers. The cloud computing providers need to make full use of the resources. With the ever increasing growing of data centers, it is very important to address the energy consumption of computer systems. Especially with the arrival of the era of cloud computing, the management of energy consumption in cloud data centers is becoming a greater challenge. For example, the Department of Energy (DOE) reports that data centers consumed 1.5% of all electricity in the US in 2006, and their power demand is growing 12% a year. Thus there are significant economic gains from research on energy saving in cloud environment.

There are already some literatures for energy-aware resource management in cloud data centers. But still some problems remain unsolved. The existing works do not take the multi-dimensional resources (such as CPU, memory and network bandwidth) into consideration when placing VM and ignore the fact that heterogeneous workloads have different resource requirements. They just deal with the placement of the new VM ignoring the optimization of the current VM allocation. As a result, these energy-aware scheduling algorithms are unpractical. The aim of proposed algorithm EST and ELMT is for managing cloud data centers to make full use of the resources. These two algorithms are based on the fact that heterogeneous workloads need a variety of resources in cloud data centers simultaneously. The optimal allocation of VMs by the two algorithms results in higher utilization of resources and reduces the power consumption.

2. Literature Review

The problem of power-aware dynamic placement of applications in virtualized heterogeneous systems as continuous optimization. At each time frame the placement of VMs is optimized to minimize power consumption and maximize performance. The authors [1] have applied a heuristic for bin packing problem with variable bin sizes and costs. They have introduced the notion of cost of VM live migration, but the information about the cost calculation is not provided. The algorithms do not handle strict SLA requirements. SLA can be violated due to variability of the workload. The work in paper [3], have a technique for minimization of power consumption in a heterogeneous cluster of computing nodes serving multiple web-applications. The technique applied to minimize power consumption is concentrating the workload to the minimum of physical nodes and switches the idle nodes off. This approach deals with the power/performance trade-off, as performance of applications degrades due to the workload consolidation. Algorithm periodically monitors the load of resources (CPU, disk storage and network interface) and makes decisions on switching nodes on/off to minimize the overall power consumption, which provides the expected performance. The algorithm runs, on a master node creates a Single Point of Failure (SPF). This may become a performance bottleneck in a large system. In addition, the reconfiguration operations are time-consuming, and the algorithm adds/removes only one node at a time, which is also a reason for slow reaction in large-scale environments. The approach, can be applied to multi-application mixed-workload environments with fixed SLAs.

The paper [9], have implemented a virtual cluster management system, allocates the resource that satisfying bandwidth guarantees. The allocation is done by a heuristic and that minimizes the total bandwidth utilized. The VM allocation is adapted, when some VMs are de-allocated. However, the VM allocation is not adapted dynamically, which depends on the current network load. Moreover, this approach does not explicitly minimize energy consumption by the network. For VM optimation author [2], have investigated the problem of energy-aware dynamic consolidation of VMs according to the current CPU utilization. However, to allow a better VM placement optimization, the VMs should be reallocated according to the current utilization of multiple system resources, including the CPU, RAM, and network bandwidth. Disk storage is usually centralized (e.g. NAS) to enable live migration of VMs and, therefore, requires specific energy-efficient management techniques. The problem arises when it comes to providing strict SLAs ensuring no performance degradation, which is required for a Cloud data center. The work proposed in paper [5], have considered the problem of energy-efficient management of homogeneous resources in Internet host. The main challenge, to determine the resource demand of each application at its current request load level and to allocate resources in the most efficient way. The system maintains an active set of servers selected to serve requests for each service. The network switches are reconfigured dynamically to change the active set of servers whenever needed. Energy consumption is reduced by switching idle servers to power saving modes (e.g. sleep, hibernation). The system deals only with the management of the CPU, but does not consider other system resources. The latency during switching nodes on/off is not taken into account.

The authors, noted that the management algorithm is fast when the workload is stable, but turns out to be relatively expensive during significant changes in the workload.The resource allocation algorithms [8], aim to accomplish the task of scheduling virtual machines on the servers residing in data centers and consequently scheduling network resources while complying with the problem constraints. Designing an energy-aware model faces performance challenges that are mostly caused by consolidation, VM migration, and server idle state configuration. The design challenges are discussed with the aim of providing a reference to be used when designing a comprehensive energy-aware resource allocation model for cloud computing data centers. In paper [6], author have two energy-aware algorithms, which often focus on only one-dimensional resources. Two algorithms are based on multiple resources (such as CPU, memory and network) that are shared by users concurrently in cloud data centres. The heterogeneous workloads have different resource consumption characteristics. Both algorithm investigate the problem of consolidating heterogeneous workloads and try to execute all Virtual Machines (VMs) with the minimum amount of Physical Machines (PMs), then power off unused physical servers to reduce power consumption. For allocating VM and VM migration [7], author have introduced techniques, in which an integrated load balance measurement was introduced. Average utilization of CPU, memory and network bandwidth during each observed period, is noted respectively. The larger the value V is, the higher the integrated utilization. The algorithm always chooses the physical machine with the lowest V and to allocate VM on it. We give this algorithm a short name V-value algorithm.

3. Proposed System

The proposed system, uses EST for Scheduling VMs and uses ELMT for Live Migration of VMs in cloud server to reduce power consumption. Fig.1 shows the EST and ELMT used in cloud Server.

Algorithm includes Three fundamental steps:

- Compute every component capability of each physical severs;
- Get every component capability of the VM;
- Get the smallest value of IUVi , and assign the VM to that sever;

The problem of VM allocation can be divided in two: the first part is the admission of new requests for VM provisioning and placing the VMs on hosts, whereas the second part is the optimization of the current VM allocation.

EST Algorithm

Energy-Aware Scheduling Technique (EST) algorithm deals with the admission of new applications for VM provisioning and places the VM on the server. In our modification, the EST algorithm, sort all VMs in decreasing order of their current CPU utilization, and allocate each VM to a host. Thus provides the least increase of power consumption and allows leveraging the heterogeneity of resources by choosing the most power-efficient nodes first.

In order to use all the provided resources in cloud data centers, heterogeneous workloads should be consolidated to the minimum number of physical machines. Thus, idle servers can be switched to the sleep mode inorder to reduce the consumption of power. The complexity of the algorithm EST is n, where n is the number of physical machines.

Algorithm 1 : EST

1) Initialize the cloud server;
2) Lets Assume n = 1and IUV = 1 where n belongs to N;
3) Compute every component Capability of server and VM;
4) Calculate IUV;
5) If Calculated IUV is less than assigned IUV, then n=N;
6) If Calculated IUV equals assigned IUV, then end;
7) If IUV = 1, Randomly turn on a new physical server r and *n* which is in sleep mode to active mode and place the VM to the sever;
8) Allocation of *vm* to server is done randomly;
9) Thus power consumption is reduced by Optimal allocation of *vm* to servers.

And the second parts of VM allocation are the optimization of the current VM allocation and dealing with where to migrate VMs. We can dynamically migrate running VMs using ELMT from underused physical machines to others, which are almost fully used and reduce consumption of power.

ELMT Algorithm

Energy-Aware Live Migration Technique (ELMT) algorithm is for optimization of the current VM allocation. It is carried out in two steps. At the first step, the algorithm ELMT chooses the VMs which are needed to be migrated. To determine when and which VMs should be migrated, we introduce node utilization threshold vector policies. The basic idea is to set upper and lower utilization thresholds for physical machine and keep the total utilization of all the VMs allocated to the physical machine between these thresholds. If the utilization of a physical machine falls below the lower threshold, all VMs have to be migrated from one physical machine to other and has to be switched to the sleep mode in order to eliminate the idle power consumption. If the utilization exceeds the upper threshold, some VMs have to be migrated from the physical machine to reduce the utilization. The aim is to preserve free resources in order to prevent SLA violations due to the consolidation in cases when the utilization by VMs increases. The difference between the old and new placements forms a set of VMs that have to be reallocated. The new placement is achieved using live migration of VMs.

If the utilization of CPU, memory and network of a single physical machine is below the given node utilization threshold, all the VMs on the physical machine have to be migrated to other physical machines. The chosen VMs are allocated to other physical machines using the EST algorithm. After all the VMs migrated to other machines, the physical machine has to be switched to the sleep mode to reduce consumption of power. The complexity of the algorithm ELMT is n*m, where n is the number of physical machines and m is the number of VMs.

Algorithm 2 : ELMT

1) Initialize the cloud server;
2) Compute every component Capability of server n;
3) Calculate node utilization threshold vector;
4) If n is less than node utilization threshold vector, then migrate n to N;
5) Turn *nth* server to sleep mode;
6) The sleep mode servers are unused
7) Thus power consumption is reduced by migrating *vm* from underused servers to those which are mostly fully used.

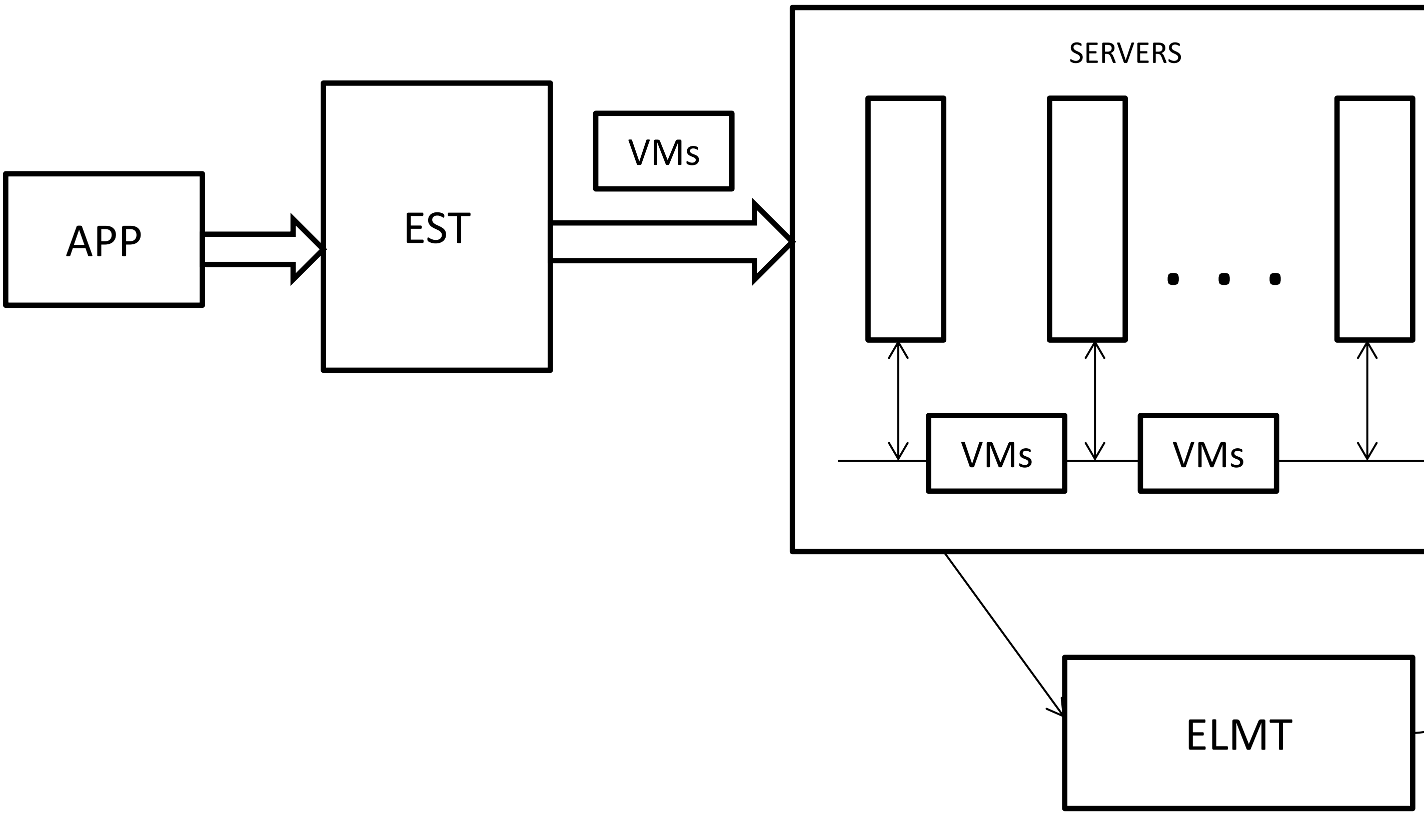

Figure 1. EST and ELMT

4. SystemArchitecture

The cloud system in our model is assumed to be deployed in a data centre which is composed by different hardware. With virtualization solutions, the users can use the resources in the form of leases safely. The IaaS cloud system uses virtualization technology to manage its resources and offers VMs to its customers is shown in Fig.2.

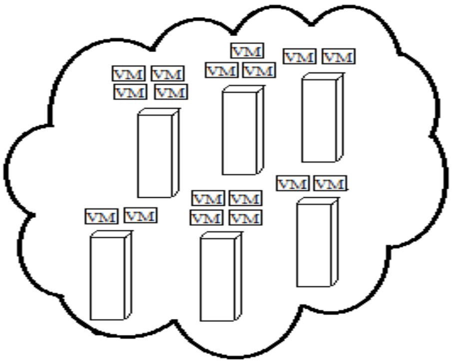

Figure 2. The Iaas Cloud Model

We assume that the IaaS cloud system consists of a set N of n physical machines. N can be represented by N = {1, 2,..., n}. Each physical machine can be multiple compositions of CPU, memory and network cards.

These heterogeneous physical machines in cloud data centers have different capabilities and speeds and can be switched on or off independently. Using virtualization technology, a set M of m virtual machines run on each physical machine. M can be represented by M = {vm1, vm2,.., vmm} . The virtual machines on a physical machine can be restarted, paused and migrated to other physical machines in the cloud data centre.

Because of the heterogeneity, we carefully consider the hardware resource heterogeneity (including CPU, memory, and network cards) when placing VMs. The power consumption of physical machines in cloud data centers is mostly determined by processor, memory, disk storage and network interface controllers.

The computing capability of server can be measured by Millions of Instructions Per Second (MIPS). The power consumption can be calculated using following formula,

$$P(u) = k \cdot P_{\max} + (1-k) \cdot P_{\max} \cdot u$$

Where Pmax is the power consumption at the peak load and k is the fraction of power consumed by the sleep server. u is the utilization. An integrated load imbalance value (ILBi) of sever i is defined as:

$$\frac{\left(\mathrm{Avg}_i - \mathrm{CPU}_u^A\right)^2 + \left(\mathrm{Avg}_i - \mathrm{MEM}_u^A\right)^2 + \left(\mathrm{Avg}_i - \mathrm{NET}_u^A\right)^2}{3}$$

Where

$$\mathrm{Avg}_i = \left(\mathrm{CPU}_i^U + \mathrm{MEM}_i^U + \mathrm{NET}_i^U\right)/3$$

Likewise calculate values for each component capability and node threshold vector to save power consumption while VM allocation and Migration. Fig.3 Shows the proposed system architecture.

International Journal of Computer Science and Engineering Communications
Vol.3, Issue 3, 2015, Page.1141-1148
ISSN: 2347–8586
www.scientistlink.org

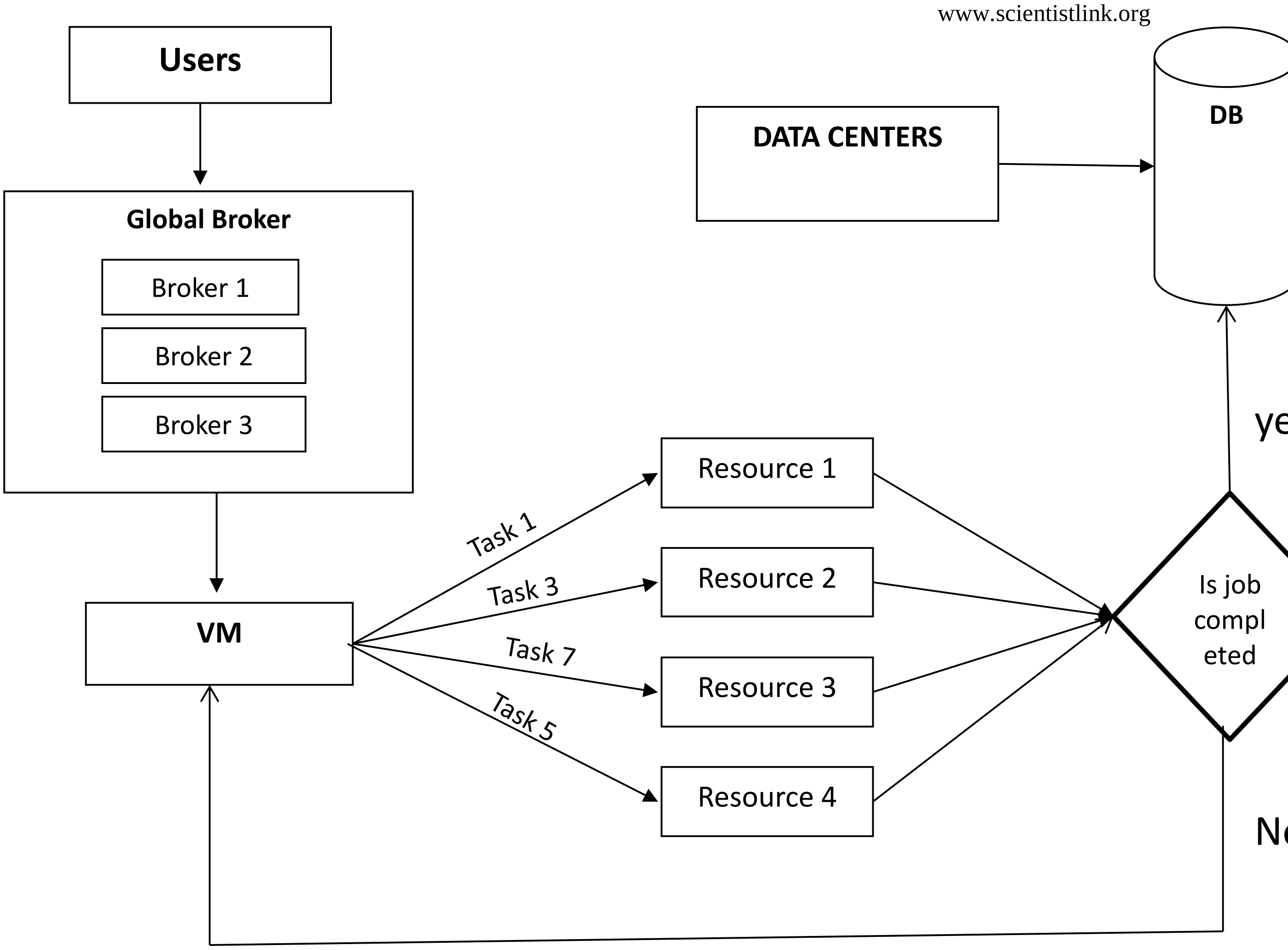

Figure 3. Proposed System Architecture

5. Scheduling Process

The scheduling process with the two proposed algorithms are as follows,

1. A user requests resources in a cloud data centre.
2. The cloud data centre provides the required resources in the form of VMs. The resource scheduling centre in the cloud data centre allocates the VM to a dedicated physical server using the algorithm EST.
3. Since the resource scheduling centre does not know when applications arrive and every VM has a life cycle, it is a dynamic scheduling problem. For a certain period of time, the resource scheduling centre uses the algorithm ELMT to migrate running VMs from underused physical servers to those which are fully used.

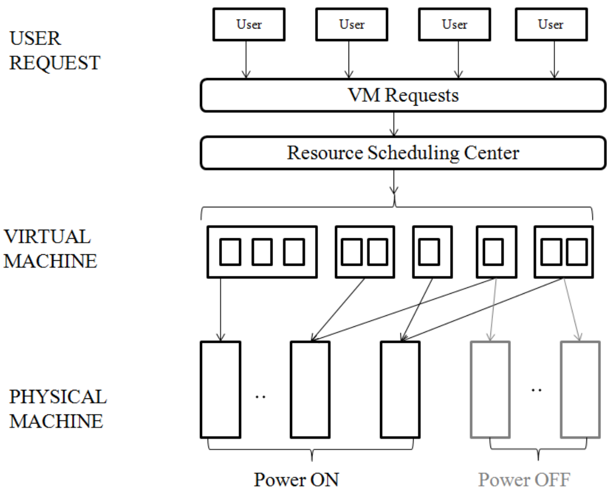

Figure 4. Scheduling Process

6. Results and Discussion

In this section, the proposed system results have been discussed. We choose CloudSim toolkit as a simulation platform. CloudSim is a modern simulation framework aimed at cloud computing environments. Table 1 Shows Simulation parameters for VM creation.

Table 1: Simulation Parameters for VM Creation

PARAMETERS	SPECIFICATION
Image Size	1000
Ram Size	512
MIPS	250
Band Width	100
PES Number	1

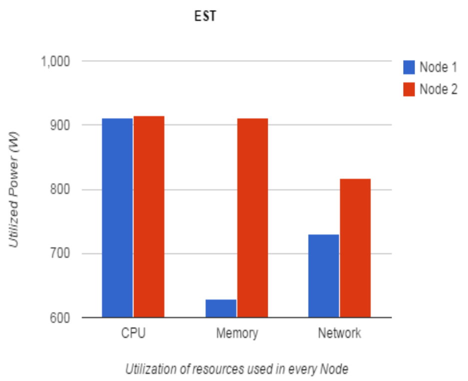

Figure 5. Utilized Power Using EST Algorithm

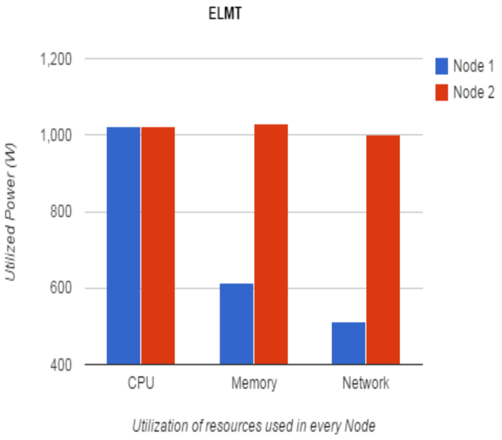

Figure 6. Utilized Power Using ELMT Algorithm

Table 2. Comparison Table

Type	Node	CPU	Memory	Network
EST	1	912	429	430
	2	915	912	817
ELMT	1	1024	512	512
	2	1024	1030	1000

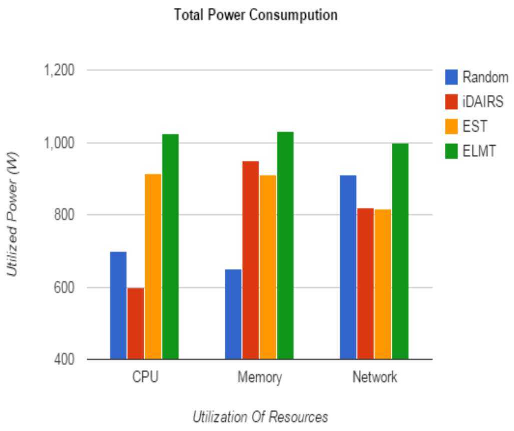

Figure 7. Comparison Of Proposed Algorithms with Standard Benchmarks

The various parameters which are given as input for setting up this environment are image size, ram size, MIPS, Band width, PES Number. We provide extensive simulations values by comparing the two algorithms EST and ELMT in Table 2. The graph of both algorithm is shown in Fig. 5 & 6. The simulation results show that the two algorithms EST and ELMT using workload- aware consolidation technique have good balanced utilizations among components in a physical sever. The optimal allocation of VMs leads to higher utilization of resources in cloud data centre. When comparing, result of ELMT shows ~0.1% increased power utilization with that of EST. Now, let us Compare EST and ELMT With Standard Benchmark algorithms to highlight its efficiency. Both the algorithms are ~ 0.12% increased power consumption when compared with Standard benchmark like iDAIRS, Random etc,. It has good balanced utilization of resources considering CPU, Memory and Network. Performance is highly efficient to that of Standard benchmark.

7. Conclusion

We proposed two algorithms to schedule a set of VMs to a set of physical machines using workload-aware consolidation technique in a cloud data centre. The goal of the two algorithms was to reduce power consumption by considering the fact that heterogeneous workloads have different resource consumption characteristics. In order to compare with different scheduling algorithms, measurements such as imbalance utilization value, average utilization of resources of a cloud data centre are developed.

8. Future Work

In Future, energy efficiency of VMs can be improved by considering additional parameters into account which may yield better performance and the algorithm may be implemented in a real cloud environment.

References

[1] A. Verma, P. Ahuja, A. Neogi, (2008) " pMapper : power and migration cost aware application placement in virtualized systems" in: Proceedings of the 9th ACM / IFIP / USENIX International Conference on Middleware, Springer, pp. 243–264.

[2] Anton Beloglazov, Rajkumar Buyya, (2012), "Energy Efficient Resource Management in Virtualized Cloud Data Centers", 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 826-831.

[3] C. Guo, G. Lu, H. Wang, S. Yang, C. Kong, P. Sun, W. Wu, Y. Zhang, (2010), "Secondnet: a data center network virtualization architecture with bandwidth guarantees", Proceedings of the CoNEXT 2010, Philadelphia, USA.

[4] Christoph Mobius, Waltenegus Dargie, Senior Member, IEEE, and Alexander Schill (2014), "Power Consumption Estimation Models for Processors, Virtual Machines, and Servers." pp. 1600-1614.

[5] E. Pinheiro, R. Bianchini, E.V. Carrera, T. Heath,(2010), "Load balancing and unbalancing for power and performancee in cluster-based systems", in: Proceedings of the Workshop on Compilers and Operating Systems for Low Power, pp. 182–195.

[6] J.S. Chase, D.C. Anderson, P.N. Thakar, A.M. Vahdat, R.P. Doyle, (2011) Managing energy and server resources in hosting centers, in: Proceedings of the 18th ACM Symposium on Operating Systems Principles, ACM, New York, NY, USA, pp. 103–116.

[7] M. Abu Sharkh, A. Ouda, and A. Shami, (2013) "Resource Allocation in a Network-Based Cloud Computing Environment: Design Challenges", Nov 2013 IEEE, pp. 46-52.

[8] TIAN Wenhong, ZHAO Yong, ZHONG Yuanliang, et al. (2011), "Dynamic and Integrated Load-Balancing Scheduling Algorithm for Cloud Data Centers". China Communications, pp. 117-126.

[9] Wood T, Shenoy P, Venkataramani A, et al. (2007), "Black-Box and Gray-Box Strategies for Virtual Machine Migration", Proceedings of the 4th USENIX Symposium on Networked Systems Design Implementation (NSDI'07),Cambridge, USA, pp. 229-242.

[10] Beloglazov A, Abawajy J, Buyya R. "Energy- Aware resource allocation heuristics for efficient Management of data centres for cloud Computing". Future generation computer systems, 2012, pp. 755-768.