iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** VI **Month of publication:** June 2016

DOI:

Volume 4 Issue VI, June 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Power Aware Mechanism for Virtual Machine Machine Provisioning and Allocation in Clouds

Malarvizhi.D[1], Dr.Venkatesa kumar.V[2]
[1]*PG Scholar, Department of CSE, Anna University Regional Campus, Coimbatore, Tamilnadu, India*
[2]*Assistant Professor, Department of CSE, Anna University Regional Campus, Coimbatore, Tamilnadu, India*

Abstract— Cloud computing is evolving technology in which pool of resources are provided as services. Resource provisioning in cloud computing achieves systematic services on client registration using services present in cloud computing. In resources provisioning there is tremendous query formation for each client for utilizing their resources i.e. memory utilization, CPU utilization, and other resources are utilizing capabilities in cloud computing. For resource provisioning in cloud two popular mechanisms are reservation and on-demand plan services. The proposed system power aware data centre allocation offers all the features that is provided by the Cloud Simulation including the features are: (i) provides support for simulation of virtualized cloud based data centre environments with dynamic VM provisioning (ii) prediction of future workload (iii) chip level power aware VM allocation (iv) random workload pattern generation (v) uses real time power saving states.
Keywords— PowerAware, Virtual Machine, Cloud Scheduling, Data Center,Performance

I. INTRODUCTION

Cloud computing has revolutionized the computing world and will continue to do so for many years to come. The core technology behind cloud computing is virtualization. Large datacenters offer computing resources in terms of virtual machines where users remotely connect to perform their computing tasks or deploy their applications. Cloud services include virtual machines, virtual platforms, and cloud-based software. Big corporations that already possess large datacenters are able to generate additional revenues (e.g., Microsoft Windows Azure, Amazon EC2). Many new companies solely based on cloud computing products are being added to the market (e.g., Salesforce.com, Rack space Hosting). On the other hand, users, especially small and medium enterprises, now have a wide array of options for securing their computing resources - they can rely on cloud resources at any level (infrastructure, platform, or software). This enables users to balance their financial requirement for computing - they can replace the upfront cost of procurement and ongoing cost of in-house maintenance with utilizing cloud resources depending on their financial goals and limitations.

Cloud computing as collection of resources (servers in datacentre), which are interconnected with each other and using virtualization technology can be scaled and adapted dynamically. Cloud computing provides customers, to start their business without purchasing any physical hardware, whereas service providers can rent their resources to customers and make their profit. Customers have the opportunity to scale up or down, the resources dynamically to provide QOS for demand varying application. Cloud computing enables dynamic and flexible application provisioning using virtualization technology.

Cloud computing, resource provisioning is an important issue as it dictates how resources may be allotted to an cloud application such that service level agreements (SLAs) of applications are met. This in turn is used to develop a heuristic algorithm, it defines an allocation scheme and it requires small number of servers. In responding to the effectiveness of the algorithm specification was evaluated in range of operating conditions. A new modification called randomly dependent priority. It is originated to have the best performance in terms of required number of servers.

Cloud computing is evolving technology in which pool of resources are provided as services. Resource provisioning in cloud computing achieves systematic services on client registration using services present in cloud computing. In resources provisioning there is tremendous query formation for each client for utilizing their resources i.e. memory utilization, CPU utilization, and other resources are utilizing capabilities in cloud computing. For resource provisioning in cloud two popular mechanisms are reservation and on-demand plan services. The proposed system PACS offers all the features that is provided by the Cloud Simulation including the features are: (i) Provides support for simulation of virtualized cloud based datacenter environments with dynamic VM provisioning (ii) Prediction of future workload (iii) chip level power aware VM allocation (iv) Random workload pattern generation (v) Uses real time power saving states.

The main contributions of this proposed system are: (i) implementation of power aware VM allocation policies like bestfit, worstfit, firstfit, binpacking. (ii) implementation of different VM schedulers like chip aware, non-chip aware, dynamic, binpacking vm schedulers, (iii) implementation of workload generation algorithms

Cloud providers face many decision problems when offering IaaS to their customers. One of the major decision problems is how to provision and allocate VM instances. Cloud providers provision their resources either statically or dynamically, and then allocate them in the form of VM instances to their customers. In the case of static provisioning, the cloud provider pre-provisions a set of VM instances without considering the current demand from the users, while in the case of dynamic provisioning, the cloud provider provisions the resources by taking into account the current users' demand. Due to the variable load demand, dynamic provisioning leads to a more efficient resource utilization and ultimately to higher revenues for the cloud provider.

II. RELATED WORK

Several researchers investigated various resource allocation problems in clouds by employing game theory. The existing formulated the resource allocation problem as a task scheduling problem with QoS constraints. They game-theoretic approximated solution. However, there is an assumption that the cloud provider knows the execution time of each subtask, which is unrealistic in cloud environments. Another existing truthful-in-expectation mechanism for resource allocation in clouds where only one type of resource was considered. A stochastic mechanism to allocate resources among selfish VMs in a non-cooperative cloud environment.

System heterogeneity plays an important role in determining the dynamics of truthful mechanisms. Our proposed mechanisms take into account the heterogeneity of the systems and that of user requests when making allocation decisions. The service provisioning problem as a generalized Nash game and proved the existence of equilibria for such game. In their model, the objective of the SaaS is to maximize its revenue satisfying the service level agreement, while the objective of the IaaS is to maximize the profit by determining the spot instances price. A problem as a Stackelberg game, and computed the equilibrium price and allocation strategy by solving the associated optimization problem. However, both studies considered only one type of VM instances, thus, the problem they solved is a one dimensional provisioning problem.

They a combinatorial auction based mechanism, CA-GREEDY, to allocate VM instances in clouds. They showed that CA-GREEDY can efficiently allocate VM instances in clouds generating higher revenue than the currently used fixed price mechanisms. However, CA-GREEDY requires that the VMs are provisioned in advance, that is, it requires static provisioning. They extended their work to dynamic scenarios by proposing a mechanism called CA-PROVISION. CA-PROVISION selects the set of VM instances in a dynamic fashion which reflects the market demand at the time when the mechanism is executed. However, these mechanisms do not consider several types of resources.

The existing system considered K uniform channels covering a certain region that is partitioned into small cells. This problem considers several cells available which in some sense correspond to several types of VMs in our study. However, in each cell a fixed number of uniform channels are available to be sold, whereas, in our case, each VM instance is composed of several types of heterogeneous resources. Furthermore, the mechanism incorporates a simple greedy metric for ordering the users that is based on the ratio of their bids to the number of requested channels. However, our proposed mechanisms incorporate bid density metrics that not only consider the structure of VMs (i.e., the multiple resources), but also take into account the scarcity of resources. In addition, we do not limit the number of available VMs for each type of VM, and we allow dynamic provisioning of VMs.

III. PROPOSED SYSTEM

Cloud computing has attracted significant attention due to the increasing demand for low-cost, high performance, and energy-efficient computing. In this large-scale, heterogeneous, multi-user environment of a cloud system, profit maximization for the cloud service provider (CSP) is a key objective. This paper focuses on scheduling virtual machines in a compute cluster to reduce power consumption via the technique of power aware resource allocation and scheduling technique. The proposed system focuses on green computing by introducing Power-Aware Data Scheduler, which provides right fit infrastructure for launching virtual machines onto host. The major challenge of the scheduler is to make a wise decision in transitioning state of the processor cores by exploiting various power saving states inherent in the recent microprocessor technology. This is done by dynamically predicting the utilization of the cloud data center. The authors have extended cloud sim toolkit to model power aware resource provisioning, which includes generation of dynamic workload patterns, workload prediction and adaptive provisioning, dynamic lifecycle management of random workload, and implementation of power aware allocation policies and chip aware VM scheduler. The experimental results show that the appropriate usage of different power saving states guarantees significant energy conservation in handling stochastic nature of workload without compromising the performance, both when the data center is in low as well as moderate utilization.

A. Cloud Model

The first heuristic, Single Threshold (ST), is based on the idea of setting upper utilization threshold for hosts and placing VMs while keeping the total utilization of CPU below this threshold. The aim is to preserve free resources to prevent SLA violation due to consolidation in cases when utilization by VMs increases. At each time frame all VMs are reallocated using power aware algorithm with additional condition of keeping the upper utilization threshold not violated. The new placement is achieved by live migration of VMs.

Typical data centers have hundreds to thousands of servers, many of which will share the same hardware and software configuration. We call such equivalence classes 'machine groups' and assume that this partitioning is performed by a separate offline process. The owner of a data center typically contracts (either internally or externally) to provide these resources to a set of applications (each associated with a customer), each with time-varying load and utility and a range of resource requirements and importance.

When allocating resources in the data center we seek to optimize the operator's business value for the data center: i.e., the revenue net of costs. This means assigning (portions of) the machines from discrete machine groups to the various applications as well as specifying the power for each machine, and thus restraining overall consumption. For this, we use a sophisticated model of the power saving modes available to modern servers and assume access to monitors of both power consumption and application demand. For each period, we use a myopic net revenue maximization objective:

$$\max = \sum_{a \in A} V_a - KE^{TOTAL} - H^{TOTAL} \quad (1)$$

Where Va is the value of the chosen allocation of machines to application (associated with a particular buyer) $a \in A$ K is the dollar cost of a kW-hour of energy, ETOTAL is the total energy used to establish and maintain the chosen allocation for the current period, and HTOTAL is the dollar cost for the hardware. The objective is thus quite straight forward the complexity comes from the constraints. Here begin by defining the buyer value, Va, i.e. the value associated with application an of some buyer.

B. Cloud Scheduling

Power aware is one of the efficient scheduling technique that utilize the principle of time slices. Here the time is divided into multiple slices and each node is given a particular time slice or time interval i.e. it utilizes the principle of time scheduling. Each node is given a quantum and its operation. The resources of the service provider are provided to the requesting client on the basis of time slice.

```
Power_aware_Load_Balancing ()
   {
Initialize all the VM allocation status to AVAILABLE in the VM state list;
Initialize hash map with no entries;
While(new request are received by the Data Centre Controller)
   Do
   {
Data Centre Controller queue the requests;
Data Centre Controller removes a request from the beginning of the queue;
If(hash map contain any entry of a VM corresponding to the current requesting user base && VM allocation status == AVAILABLE)
   {
 The VM is reallocated to the user base request;
    }
Else
          {
 Allocate a VM to the user base request using Round Robin Algorithm;
Update the entry of the user base and the VM in the hash map and the VM state list;
    }
   }
```

Power aware algorithm focuses on the fairness. Power aware uses the ring as its queue to store jobs. Each job in a queue has the same execution time and it will be executed in turn. If a job can't be completed during its turn, it will be stored back to the queue waiting for the next turn. The advantage of Power aware algorithm is that each job will be executed in turn and they don't have to be waited for the previous one to get completed. But if the load is found to be heavy, Power aware will take a long time to complete all the jobs. The Cloud Sim toolkit supports Power aware scheduling strategy for internal scheduling of jobs.

C. Virtual Machine Provisioning

Moving the contents of a VM's memory from one physical host to another can be approached in any number of ways. However, when a VM is running a live service it is important that this transfer occurs in a manner that balances the requirements of minimizing both downtime and total migration time. The former is the period during which the service is unavailable due to there being no currently executing instance of the VM; this period will be directly visible to clients of the VM as service interruption. The latter is the duration between when migration is initiated and when the original VM may be finally discarded and, hence, the source host may potentially be taken down for maintenance, upgrade or repair. It is easiest to consider the trade-offs between these requirements by generalizing memory transfer into three phases:

1) *Push phase*: The source VM continues running while certain pages are pushed across the network to the new destination. To ensure consistency, pages modified during this process must be re-sent.
2) *Stop-and-copy phase:* The source VM is stopped, pages are copied across to the destination VM, then the new VM is started.
3) *Pull phase:* The new VM executes and, if it accesses a page that has not yet been copied, this page is faulted in ("pulled") across the network from the source VM.

D. Power aware Allocation Policy

The introduce factors to model the power consumption of single physical machine. Power consumption (Watts) of a physical machine is sum of total power of all components in the machine. estimated power consumption of a typical server (with 2x CPU, 4x memory, 1x hard disk drive, 2x PCI slots, 1x mainboard, 1x fan) in peak power (Watts) spends on main components such as CPU (38%), memory (17%), hard disk drive (6%), PCI slots (23%), mainboard (12%), fan (5%). that power consumption of a physical machine (P(.)) is linear relationship between power and resource utilization (e.g. CPU utilization). The total power consumption of a single physical server (P(.)) is:

$$P(U_{cpu}) = P_{idle} + (P_{max} - P_{idle})U_{cpu}$$

$$U_{cpu}(t) = \sum_{c=1}^{PE_j} \sum_{i \in r_j(t)} \frac{minps_{i,c}}{MIPS_{j,c}}$$

In which:

Ucpu(t): CPU utilization of the physical machine at time t, $0 \leq$ Ucpu(t) ≤ 1
Pidle: the power consumption (Watt) of the physical machine in idle, e.g. 0% CPU utilization
Pmax: the maximum power consumption (Watt) of the physical machine in full load, e.g. 100% CPU utilization
mipsi,c: requested MIPS of the c-th processing element (PE) of the VMi
MIPSj,c: Total MIPS of the c-th processing element (PE) on the physical machine Mj

The number of MIPS that a virtual machine requests can be changed by its running application. Therefore, the utilization of the machine may also change over time due to application. The link the utilization with the time t. We re-write the total power consumption of a single physical server (P(.)) with Ucpu(t) as:

$$P\left(U_{cpu}(t)\right) = P_{idle} + (P_{max} - P_{idle})U_{cpu}(t)$$

and total energy consumption of the physical machine (E) in period time [t0, t1] is defined by:

$$E = \int_{t0}^{t1} P\left(U_{cpu}(t)\right) dt$$

Dynamic resource placement and provisioning are useful techniques for handling the multi-time-scales variations in the capacity of resources. Supported by making decision algorithms, dynamic resources allocation is used to perform consolidation targeting goals such as minimizing total power consumption, without substantial degradation on performance.

IV. EXPERIMENTAL RESULTS

The data center that comprises 10 heterogeneous physical nodes. Each node is modeled to have one CPU core with performance equivalent to 2000, 2500, 3000 or 3500 Million Instructions Per Second (MIPS), 16 GB of RAM, 10 GB/s network bandwidth and 1 TB of storage. Power consumption by the hosts is defined by the model. According to this model, a host consumes from 175 W with 0% CPU utilization up to 250 W with 100% CPU utilization. Each VM requires one CPU core with maximum of 1000, 2000, 2500 or 3250 MIPS, 1 GB of RAM, 100 Mb/s network bandwidth and 1 GB of storage. However, during the lifetime VMs may use fewer resources creating the opportunity for dynamic consolidation. The CPU MIPS ratings are equivalent to Amazon EC2 instance types. The users submit requests for provisioning of 500 heterogeneous VMs. Each VM is randomly assigned a workload trace from one of the servers from the workload data. Initially, VMs are allocated according to their parameters assuming 100% utilization.

TABLE I
RESOURCE UTILIZATION FOR CLOUD PROVIDER

Techniques	Time								
	1	5	10	15	20	25	30	35	40
G-VMPAC	9.4	10.4	10.8	11.3	11.6	11.6	11.8	11.8	11.8
Power Aware	8	8.5	9.2	9.6	9.9	10.1	10.4	10.5	10.5

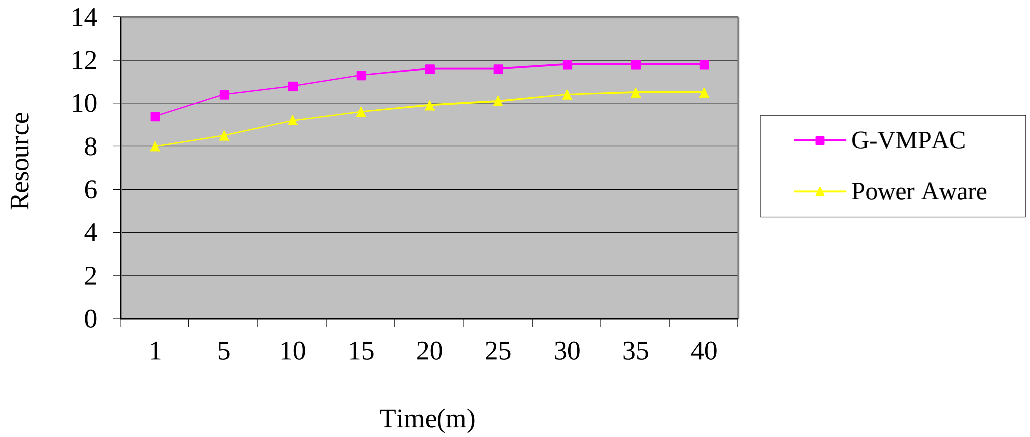


Fig. 1 Comparison of resource utilization existing with proposed system

TABLE III
POWER CONSUMPTION FOR CLOUD PROVIDER

Techniques	Time								
	1	5	10	15	20	25	30	35	40
G-VMPAC	17	21	27	34	45	51	59	75	89
Power Aware	5	8	10	22	33	43	52	68	71

Volume 4 Issue VI, June 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

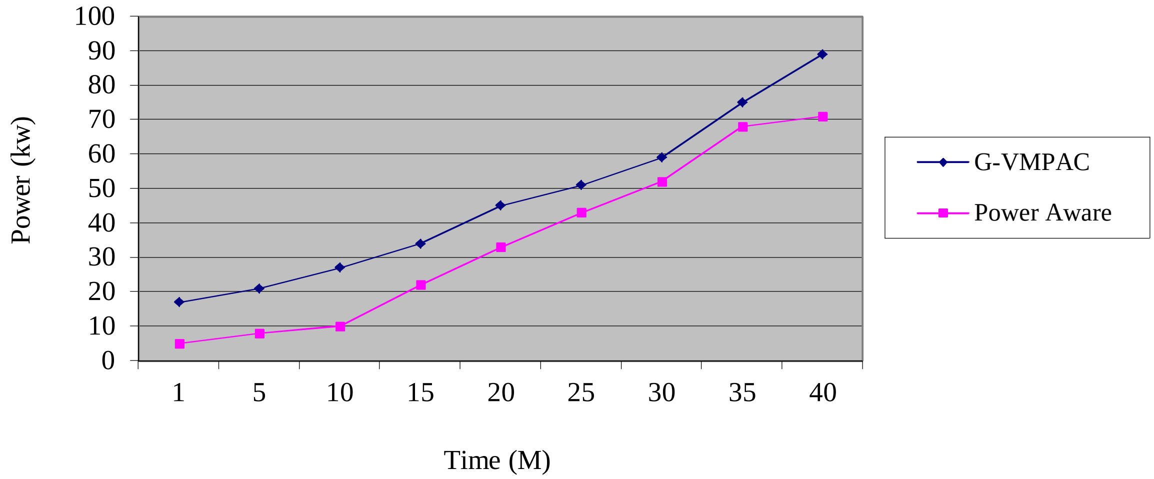


Fig. 2 Comparison of powers utilization existing with proposed system

TABLE IIIII
VM ENERGY CONSUMPTION RATE FOR CLOUD PROVIDER

Techniques	Time								
	1	5	10	15	20	25	30	35	40
G-VMPAC	0.23	0.27	0.31	0.36	0.42	0.53	0.57	0.63	0.67
Power Aware	0.14	0.19	0.23	0.27	0.31	0.38	0.43	0.45	0.54

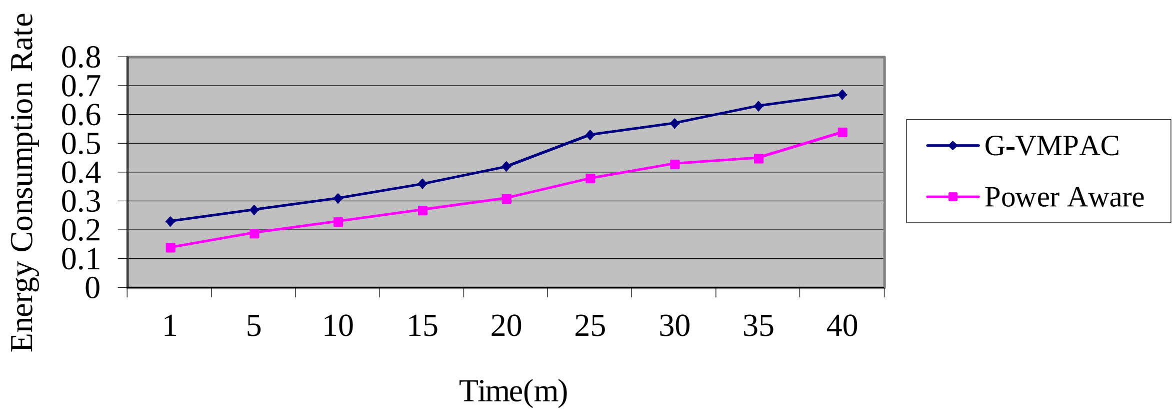


Fig 3 Comparison of energy utilization existing with proposed system

TABLE IVV
EXECUTION TIME FOR CUSTOMERS

Techniques	Cloudlet								
	1	5	10	15	20	25	30	35	40
G-VMPAC	253	286	298	318	397	432	497	535	588
Power Aware	150	185	210	245	298	345	387	430	550

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

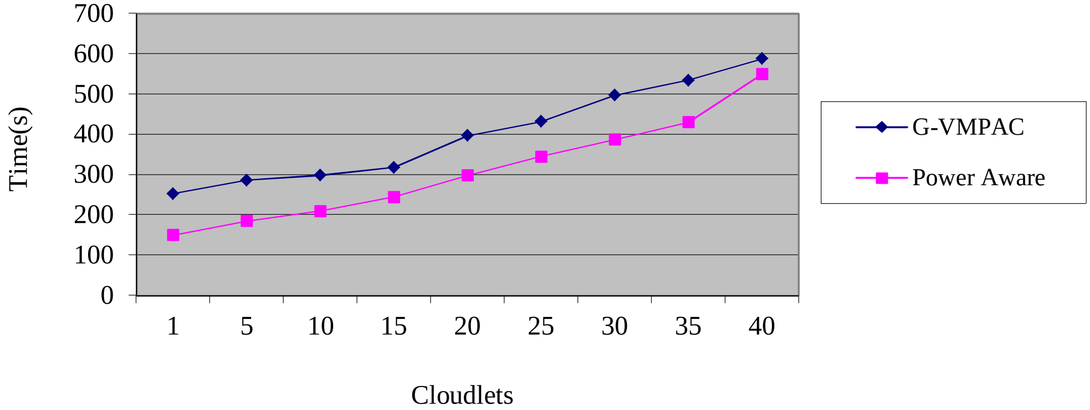


Fig. 4 Comparison of cloud job finishing time existing with proposed system

V. CONCLUSIONS

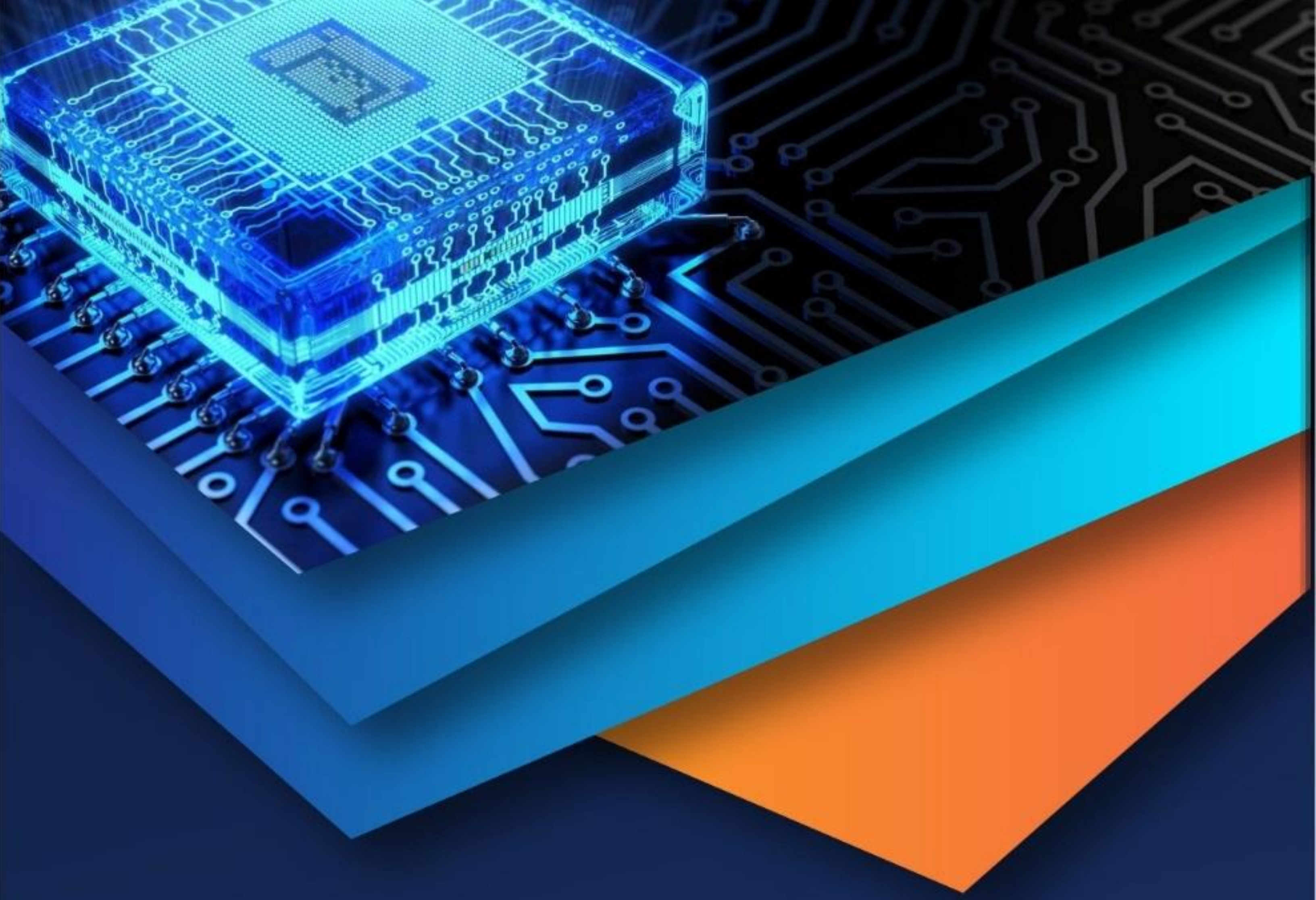
Power aware algorithm is a simulation framework that allows seamless modeling, simulation and experimentation of emerging Cloud computing infrastructure and power aware services. Power aware algorithm uses the CloudSim core simulation engine which is a discrete event driven simulator. Power aware algorithm is a continuous event driven simulator comes with an efficient framework for managing datacenters using datacenter manager and host using host manager. Power aware algorithm offers all the features that is provided by the CloudSim including the additional features: (i) provides support for simulation of virtualized cloud based datacenter environments with dynamic VM provisioning (ii) prediction of future workload (iii) chip level power aware vm allocation (iv) random workload pattern generation (v) uses real time power saving states The main contributions of this paper are: (i) implementation of different vm allocation policies like bestfit, worstfit, firstfit, binpacking etc. (ii) implementation of different vm schedulers like chip aware, non-chip aware, dynamic vm schedulers, (iii) implementation of workload generation algorithms. The unique features of Power aware algorithm are: (i) it automatically take care of power saving of datacenter by using power states (ii) predicting the future workload and kind of incoming jobs By using Power aware algorithm, researchers and industry-based developers can test the performance of a newly developed application service in a controlled and easy to set-up environment.

For our future work, we would like to introduce an optimization policy to meet the cost requirement. Secondly, a test bed can be created to investigate the algorithm behavior with multiple numbers of resources. Thirdly, we would also investigate this technique on real cloud setup and check what will be its exact reaction of on environment. This can be a small social step for significant decrease in emission of carbon dioxide along with reduction in infrastructure and operating cost.

REFERENCES

[1] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, Algorithmic Game Theory. Cambridge Univ. Press, 2007.
[2] G. Wei, A. Vasilakos, Y. Zheng, and N. Xiong, “A Game-Theoretic Method of Fair Resource Allocation for Cloud Computing Services,” The J. Supercomputing, vol. 54, no. 2, pp. 252-269, 2010.
[3] N. Jain, I. Menache, J. Naor, and J. Yaniv, “A Truthful Mechanism for Value-Based Scheduling in Cloud Computing,” Theory of Computing Systems, pp. 1-19, http://dx.doi.org/10.1007/s00224-013-9449-0, 2013.
[4] Z. Kong, C.-Z. Xu, and M. Guo, “Mechanism Design for Stochastic Virtual Resource Allocation in Non-Cooperative Cloud Systems,” Proc. IEEE Fourth Int’l Conf. Cloud Computing, pp. 614-621, 2011.
[5] Y. Wang, A. Nakao, and A.V. Vasilakos, “Heterogeneity Playing Key Role: Modeling and Analyzing the Dynamics of Incentive Mechanisms in Autonomous Networks,” ACM Trans. Autonomous and Adaptive Systems, vol. 7, no. 3, article 31, 2012.
[6] D. Ardagna, B. Panicucci, and M. Passacantando, “Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems,” IEEE Trans. Services Computing, vol. 6, no. 4, pp. 429-442, Oct.-Dec. 2013.
[7] V. Di Valerio, V. Cardellini, and F. Lo Presti, “Optimal Pricing and Service Provisioning Strategies in Cloud Systems: A Stackelberg Game Approach,” Proc. IEEE Sixth Int’l Conf. Cloud Computing, pp. 115-122, 2013.
[8] X. Zhou, S. Gandhi, S. Suri, and H. Zheng, “ebay in the Sky: Strategy-Proof Wireless Spectrum Auctions,” Proc. ACM Mobi-Com, pp. 2-13, 2008.
[9] X. Zhou and H. Zheng, “Trust: A General Framework for Truthful Double Spectrum Auctions,” Proc. IEEE INFOCOM, pp. 999-1007, 2009.
[10] G.S. Kasbekar and S. Sarkar, “Spectrum Auction Framework for Access Allocation in Cognitive Radio Networks,” IEEE/ACM Trans. Networking, vol. 18, no. 6, pp. 1841-1854, Dec. 2010.
[11] F. Wu and N. Vaidya, “A Strategy-Proof Radio Spectrum Auction Mechanism in Noncooperative Wireless Networks,” IEEE Trans. Mobile Computing, vol.

12, no. 5, pp. 885-894, May 2013.
[12] J. Jia, Q. Zhang, Q. Zhang, and M. Liu, “Revenue Generation for Truthful Spectrum Auction in Dynamic Spectrum Access,” Proc. ACM MobiHoc, pp. 3-12, 2009.
[13] S. Zaman and D. Grosu, “Combinatorial Auction-Based Dynamic VM Provisioning and Allocation in Clouds,” Proc. IEEE Third Int’l Conf. Cloud Computing Technology and Science, pp. 107-114, 2011.
[14] S. Zaman and D. Grosu, “Combinatorial Auction-Based Allocation of Virtual Machine Instances in Clouds,” J. Parallel and Distributed Computing, vol. 73, no. 4, pp. 495-508, 2013.
[15] H. Kellerer, U. Pferschy, and D. Pisinger, Knapsack Problems. Springer, 2004.
[16] A. Mu’alem and N. Nisan, “Truthful Approximation Mechanisms for Restricted Combinatorial Auctions,” Proc. 18th Nat’l Conf. Artificial Intelligence, pp. 379-384, 2002.
[17] D. Lehmann, L. O_callaghan, and Y. Shoham, “Truth Revelation in Approximately Efficient Combinatorial Auctions,” J. ACM, vol. 49, no. 5, pp. 577-602, 2002.
[18] M.M. Nejad, L. Mashayekhy, and D. Grosu, “A Family of Truthful Greedy Mechanisms for Dynamic Virtual Machine Provisioning and Allocation in Clouds,” Proc. IEEE Sixth Int’l Conf. Cloud Computing, pp. 188-195, 2013.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY