Energy-Aware Scheduling Framework for resource allocation in a virtualized cloud data centre

Kamali Gupta [1], Vijay Katiyar [2]

[1]Department of Computer Science & Engineering, Maharishi Markandeshwar University, Ambala, India
kamaligupta13@gmail.com
[2] Department of Computer Science & Engineering, Maharishi Markandeshwar University, Mullana, India
katiyarvk@mmumullana.org

Abstract- **Cloud paradigm is an embryonic computing model that in its vicinity stresses on proficient utilization of computing resources. Data centers that host and service cloud applications ingest enormous amount of energy, leading to massive emission of carbon footprints to the atmosphere and high operational expenditures. Consequently, there is a need to establish synergy between data centre resources for optimum resource utilization and strategies needs to be devised that can considerably reduce energy consumption in cloud data center. This paper elucidates an architectural framework for computation of energy spent in scheduling resources on hosts. The framework has been implemented for bin-packing techniques and explicates minutiae about broker components involved in scheduling process.**

Keyword- Cloud Computing, Energy Consumption, VM Migration, Resource Scheduling.

I. INTRODUCTION

Contemporary [1,2] resource-intensive enterprises has engendered demands for high performance computing infrastructures. Proliferation of IT services to be used by diverse range of cloud users has led to construction of large-scale energy hungry data centers that can facilitate computing services. Despite of the improvisations introduced in energy consumption models, service providers are confronted with challenges of reduction in energy consumption and CO_2 emission.

The rationale [3] in the wake of explosion of energy emission is increase in number of computer usage due to increase in number of IT practitioners. As an upshot, size of data centres has increased. Moreover, exploiting energy-aware resource provisioning to its fullest extent can subsequently provide a solution to the forefront issues. [4] Service virtualization and consolidation are acting as inherent practices that can escort energy-efficient datacenter architectures. It has effectually led to efficient resource utilization. VM provisioning can be sighted as a multidimensional bin packing problem comprising of capricious bin configurations and cost parameters. Virtualization technique embraces server consolidation process and a VM live-migration technique that has validated to be efficient in drastic reduction in energy consumption in high-performance cloud datacenters. However, I/O virtualization has excavated grounds for performance degradation posed by overheads encountered in vm migrations and needs to be addressed urgently.

The organization of paper is laid out as literature review in beginning followed by research focus. The IV section elucidates architectural framework trailed by working prototype. The last section illustrates conclusion and future work.

II. LITERATURE REVIEW

The research work presented in [5] explores method to manage data intensive distributed programming paradigms (like MapReduce and Dryad) that assists practitioners to effortlessly parallelize the processing of huge data sets. Deployment of such data intensive computing infrastructures is of significant concern due to rise in cost. The work carried out in the study dynamically adjusts the size of resource allocations to make it precise to suite parallelized tasks. The objective is to match the hardware configuration effectively with tasks requirements. This consents the system to amortize the cost endured due to idle power usage of servers transversely to the huge workloads with the aim to maximize energy efficiency and throughput.

In research work of [6], the technique of task activity vectors have been explored and deployed for characterizing applications as per resource utilization. On top of it, migration and co-scheduling policies have been applied to perk up performance and energy efficiency by coalescing applications that use complementary resources and performs frequency scaling in situation of contention owing to discouraging workloads. The experimentation has been performed on KVM virtualization environment and Linux operating system kernel. The upshot provides evidence that resource-conscious scheduling can significantly reduce the energy delay product.

The technique expounded in [7] exploits virtualization by blind consolidation of virtual machine instances to optimize power consumption of data centres. The work explicates four models, specifically, the application model, the migration model, the energy model and the target system model that identifies the performance interferences caused in utilization of resources. Besides, it computes cost occurred due to vm migrations. A consolidation fitness metric is also illustrated that can estimate the merit after consolidating vms on hosts. The proposed energy-aware scheduling algorithm minimizes the total power consumption of servers considerably.

High performance computing clouds are leaned towards reducing energy consumption. The investigated technique of [8] encompasses energy-aware allocation mechanisms along the stated two dimensions: interval times of virtual machines and multi-dimensional resources. The work elaborates that total energy consumption is equivalent to summation of total completion times of all servers. The two proposed heuristics, precisely, MinDFT-ST and MinDFT-FT algorithms minimizes the total completion times of all machines in a data center.

Virtualization technology empowers cloud infrastructures to enable construction of elastic computing environments, there by, facilitating energy and resource cost optimizations. Consolidation of resources enhances resource utilization and reduces carbon footprints considerably. Additionally, failures in highly networked computing environments can pessimistically impact system performance, prohibiting it from achieving the stated objectives. The research work of [9] proposes algorithms to dynamically construct virtual clusters to facilitate the execution of users' jobs. The algorithm leverages virtualization tools to provision proactive fault-tolerance and energy efficiency mechanism for virtual cluster. Simulations were conducted to mitigated energy inefficiencies by injecting random synthetic jobs using the most recent version of Google cloud tracelogs.

The study explained in [10] works upon three task scheduling heuristics, specifically, RASA, TPCC and PALB. The performances of the algorithms are dependent upon the parameters: cost effectiveness, power efficiency, Co2 emissions, location and climatic conditions. A check on power wastage caused due to data center cooling systems has been kept to minimize it to a considerable extent and simulations have been carried out using CloudSim tool.

The work presented in [11] explicates a precise VM placement algorithm that is oriented towards ensuring energy efficiency and adheres to Service Level Agreement. The mathematical model of the proposed algorithm is comprehended by a sophisticated data analytic system (implemented as a cloud-based solution). The exactitude of the algorithm can be corroborated as each VM can build its own data model as when required over an appropriate time horizon. In consequence, the data model reflects precisely the attributes of resource usage of the VM. The cloud-based solution algorithm is capable of communicating synchronously or asynchronously with data analytic service. The experimental setup constitutes evaluating several advanced data modeling and forecasting techniques. Experimental results elucidate reduction in energy consumption and vm migration along with SLA avoidance.

As discussed in [12] servers in data centres operates in either power-saving or active mode. The strategy is to reduce energy consumption by executing the required number of virtual machine instances in active state mode and placing rest of the vms in power-saving mode. But switching from one state to another also poses additional energy costs. The problem is devised as a boolean integer linear programming model and heuristic based approach has been used to carry out simulations for significant energy reduction.

The researcher [13] develops a novel holistic energy-aware VM scheduling framework for private clouds (Snooze) in a virtualized cloud environment. An extensive evaluation has been carried out to adjudge energy and performance implications of the proposed system on 34 power-metered machines under dynamic web workloads. The results demonstrate that the energy saving mechanisms can be allowed to dynamically scale data center energy consumption in proportion to the load.

The study of [14] presents a line of investigation on Power management strategies for enterprise servers on the basis of Dynamic Voltage and Frequency Scaling i.e. (DVFS). DVFS provisions transition of processors from high to low-power states. The processors undergoes deep sleep mode to trim down energy consumption. This mode allows the server to get configured to use Direct Memory Access (DMA) mechanism to situate arriving packets into memory buffers for enabling processing in active state. The Virtual grouping scheme is improved to handle resources with load balancing mechanism. The system is enhanced with optimization mechanism to deal with relative response time parameter. The system is espoused to serve Dual in-line Memory Module and Dynamic Random Access Memory components.

The research work proposed in [15] confronts issues pertaining to maximizing the performance of a data center taking into consideration the deadline of competing tasks. The study also addresses issues concerning reduction in power consumption and thermal constraints. Thermal aware resource allocation techniques are developed. The experimentation conducted validates approximately 17% improvement in the reward collected and 9 % reduction in power consumption in contrast to other allocation mechanisms.

III. RESEARCH FOCUS

The research study presented in the paper is directed towards deriving an architectural framework for broker component of IaaS cloud interface that can optimally establish trade-off between reduction in energy consumption and performance delivery. The key challenge that lies in such arrangement is to determine the situation which causes the migrations of Vm instances. Besides, deciding upon the vms that are to be migrated along with hosts on which migration is to be performed is a tedious task. Moreover, identifying the physical hosts that can be turned off without compromising on the service quality is also cumbersome in today's fluctuating heterogeneous service request. Hence, comprehending solution to such issues has become a dire need in present computing paradigm. Techniques needs to be implemented that can combine multi-faceted resource allocation policies developed with different objectives to extend an answer to the common problems encountered by consumers and users in cloud computing. The integration of components each capable of retorting to a different challenge can lead to an optimal solution. Hence, the research study explores and integrates these scattered components into an architectural framework that can be deployed for effective resource allocations.

IV. ARCHITECTURAL FRAMEWORK

The architectural framework is presented in Figure 1. The major components are conferred below:

Service Providers: Consumers in a typical cloud [16], is an entity that is allied with the environment in order to obtain required services. The cloud that spans across the globe is not bounded to any location, thus does its functionality is not limited due to geographical constraints. Consequently, vendors are available in assorted regions to facilitate such consumers. The vendor who offer services to the consumers or users are providers of the services as requested by the users, hence are termed as service providers. They are authorized agents and can submit the requests into cloud using some authentication mechanism. Such procedure allows only designated merchants to log in the environment for secured services over web. The service requests raised by the providers are submitted to the broker who further activated the designated module covered under its operations to enable services in a distributed manner encompassing efficient usage of available resources.

Broker: The Broker is an entity that is associated with provisioning of resources registered or available with it within constraints of service level agreements as submitted by the request generator. The entity for its efficacious operations is divided into a number of sub-entities each envisioned with own set of functionalities stated as under:-

Service Analyzer: The service Analyzer [17] as a part of Broker module interprets and scrutinizes the service requests for acceptance or rejection. Consequently, the module upholds newest load and energy information by accumulating it from VM Manager and Energy Monitor module. The rejected requests get back to the service catalogue for reconsideration.

Resource Catalogue and Profiler: The module endorses specific information pertaining to available service providers and the type of service registered under their catalogue for offering special privileges to them. It also maintains details of computing capacities and power requirements of various resources. The profiler module in consultation with the analyzer ensures prioritization in the cloud scheduling environment as per the services that can be catered by providers depending upon resource availability.

Energy Monitor & calculator: The module on the basis of energy consumption and load information identifies the physical host to be powered on/off. The idea behind switching on/off the host machines is to reduce the energy consumption in cloud data centres by turning off idle machines. The calculator modules computes total energy spent on allocation process and monitor further communicates this information to resource catalogue and scheduler for further processing.

Service Scheduler: The scheduler module executes the process of mapping of vm instances to physical host machines. The component performs resource entitlements for allocated VMs. Upon integrating in its functionality, the proposed technique, the scheduler assigns vms to allocate the physical machines on which the virtual machines are to be stimulated. The module comprises of two sub-sections, namely, vm selection and vm allocation. VM selection module selects the virtual machine instances (among the set of available resources) that needs to be migrated while the vm allocation module assigns physical hosts to these migrating vm instances.

VM Manager: The manager module ensures the availability of VMs along with information relevant to resource entitlements. The element accomplishes migration of virtual machines across physical machines in a virtualized cloud environment.

VM Consolidator: [18] The module is responsible for consolidation of vm instances on a single host. The technique is used to optimize the count of active physical hosts as per current resources utilization.
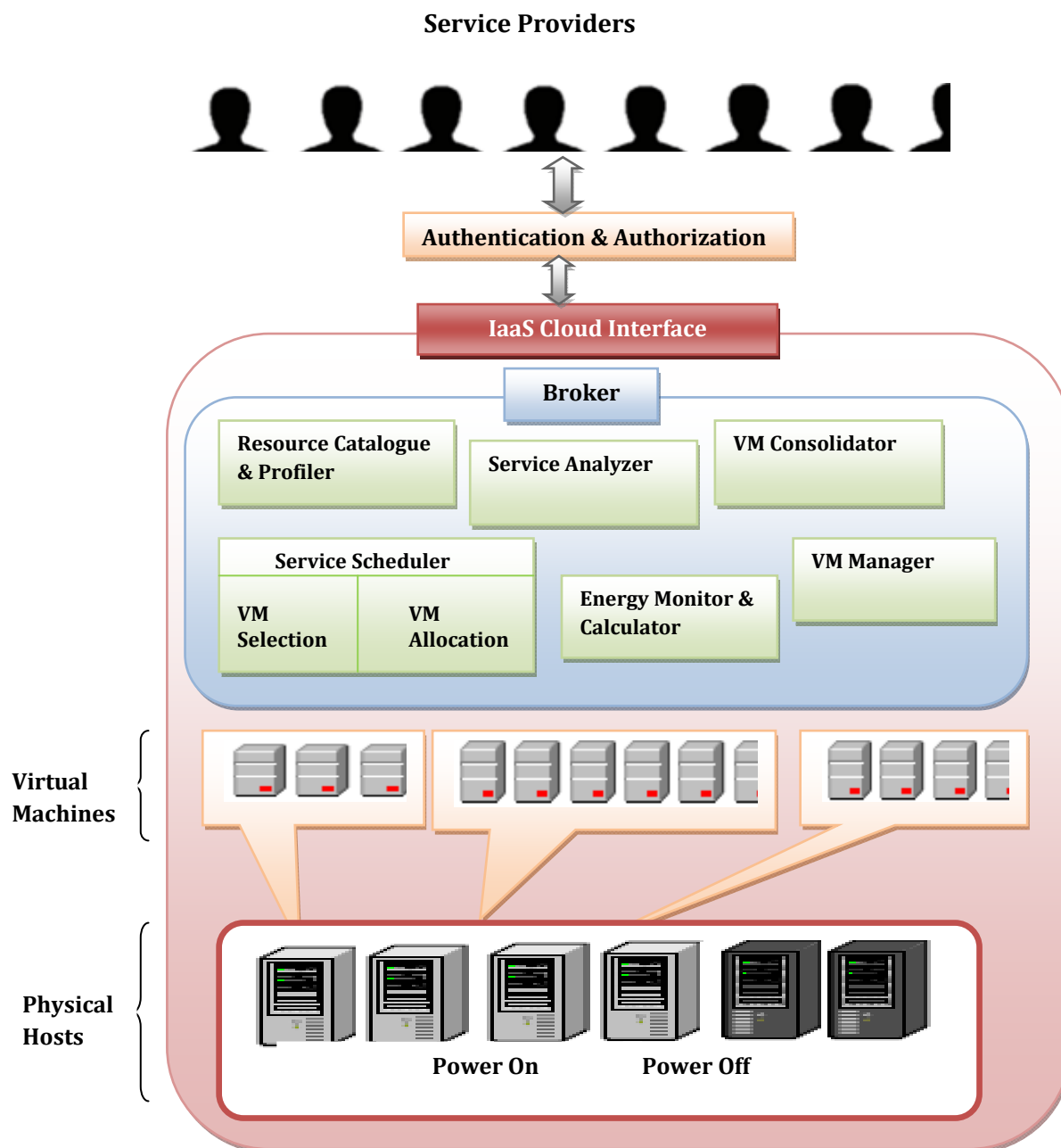
Fig.1. Architectural framework for proposed technique

V. WORKING PROTOTYPE

The framework has been developed for facilitating heuristic developed in [19] which is a bin-packing technique that is oriented towards allocation of resources (virtual machine instances) to hosts in a way that can minimize the energy consumption. Bin-Packing refers to packing of hosts machines with the available virtual machines with the objective to turn off idle servers. The technique encompasses migration of vm instances across physical machines in a data center that are capable of serving the raised requests. The architecture talks about scheduling at infrastructure level as at this level resource allocation is done. In contrast to it, task allocation is done at platform level wherein the jobs or tasks submitted by users are presented before the brokers for their execution on resources available. The technique implemented in this work is an enhancement to already existing bin-packing solutions presented in [20]. The workflow of the heuristic has been demonstrated in Figure 2.
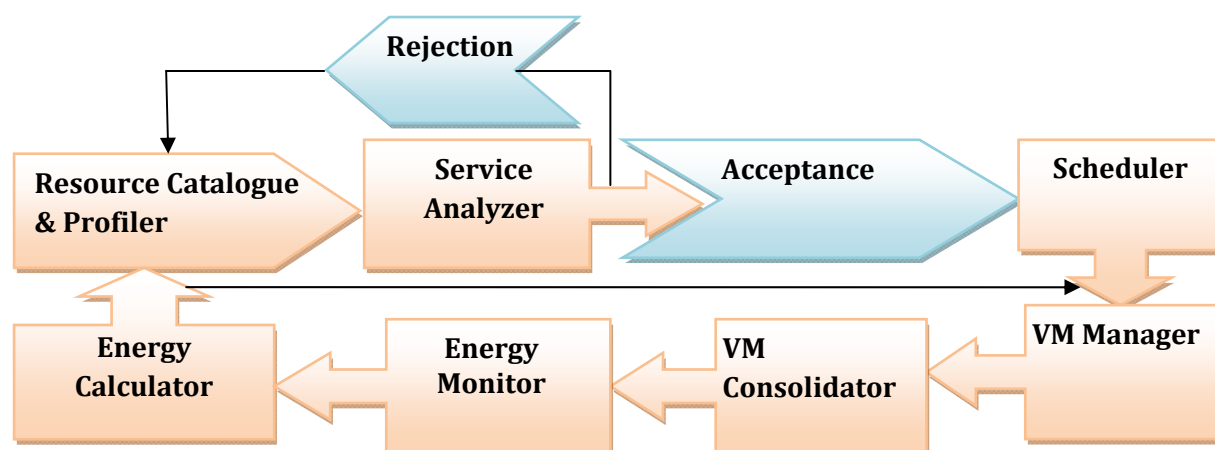
Fig.2. Workflow of Broker Module

Initially, all the servers are unassigned. The resource catalogue module initializes the vm instances and servers with their computing capacities and power requirements. The profiler filters the requests keeping in mind geographical and other constraints and passes the requests to the Analyzer component. This section scrutinizes the requests for acceptance or rejection on the basis of availability of resources. The rejected requests are reverted to Resource Catalogue for upholding into the request queue and accepted requests along with the information concerning resources are send to Scheduler. The module as per computing capacities and power requirements of servers generates a mapping schedule by assigning vm instances to servers. VM Manager migrate the vm instances across servers. The consolidator module consolidates the migrated instances onto the designated physical machines. Further, the monitor turns off the idle servers and calculator calculates the energy spent in the process as per equation:

Total Energy Spent $= \sum_{i=0}^{s} Pi(F, t)$ //summation of power consumption of all servers

Where $P_i(F,t) = 0.7\, P_{max}^i + 0.3\, P_{max}^i * U_i(F,t)$ // Models power consumed by a server P_i

$P_i(F,t)$ is power consumption of a server, $U_i(F,t)$ is Utilization of server for a placement in time t and is computed by equation:

$$U_i(F,t) = \sum_{K=1}^{V} F_{i,k} * \left(\frac{Computing\ requirement\ of\ virtual\ machine\ k\ on\ server\ i}{Computing\ capacity\ of\ server\ i} \right)$$

$F_{i,k}$ specifies hosting of vm k on server i. It can contain two values 0 or 1, if 1 then vm is hosted otherwise not. As a final step, the computer energy for a schedule and information regarding turned off hosts is communicated to Catalogue and Scheduler for further processing.

VI. CONCLUSION

Cloud Computing as a paradigm from recent years is striving hard to enlarge the services it offers to its consumers by exploring newest mechanisms and policies. The technology has gained popularity in terms of dissemination of resources as pay-per-use model that has facilitated its accessibility trouble-free. But as the number of users reaching cloud environment is escalating, the power management and CO_2 emissions are profoundly becoming foremost area of concern for service providers. Several bin-packing heuristics have been proposed and implemented to overcome this issue and focuses on packing of servers with virtual machines, consequently turning off the unused bins to save energy consumption. The paper presents architectural framework for one such proposed technique implemented in earlier studies. Various components are aptly discussed and future consideration of the researcher is to integrate the existing work with market-oriented resource management techniques that can assist in optimizing energy and resource efficiency.

REFERENCES

[1]   A. Beloglazov, R. Buyya, "Energy Efficient Resource Management in Virtualized Cloud Data Centers" In Pro. 10th IEEE/ACM Int. Conference on Cluster, Cloud and Grid Computing, pp. 826-831, 2010.
[2]   E. Arianyan, H. Taheri, S. Sharifiaan, "Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centers", ELSEVIER, Computers and Electrical Engineering, pp. 1-19, 2015.
[3]   A.P. Xiong, C.X. Xu, "Energy Efficient Multiresource Allocation of Virtual Machine Based on PSO in Cloud Data Center", Hindawi Publishing Corporation, pp.1-9, 2014.
[4]   D. Liu, N. Han, "An Energy-efficient Task Scheduler in Virtualized Cloud Platforms", International. Journal of Grid DistributionComputing, vol. 7, no.3, pp.123-134, 2014.
[5]   W. Xiong, A. Kansal ,"Energy Efficient Data Intensive Distributed Computing", IEEE Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, pp. 1-10, 2010.

[6]   A. Merkel, J. Stoess, F. Bellosa, "Resource-conscious Scheduling for Energy Efficiency on Multicore Processors", ACM EuroSys'10, 2010.
[7]   M. Sharifi, H. Salimi, M. Najafzadeh, "Power-efficient distributed scheduling of virtual machines using workload-aware consolidation techniques", Springer, J Supercomputers, July 2011.
[8]   N.Q. Hung, D.K. Le, N. Thoai, N.T. Son, "Heuristics for Energy-Aware VM Allocation in HPC Clouds", Springer International Publishing Switzerland, pp. 248-261, 2014.
[9]   A.M. Sampaio, J.G. Barbosa, "Towards high-available and energy-efficient virtual computing environments in the cloud", ELSEVIER, Future Generation Computer Systems, pp. 30-43, 2014.
[10] M. Uddin, Y. Darabidarabkhani, A. Shah, J. Memon, "Evaluating power efficient algorithms for efficiency and carbon emissions in cloud data centers:A review", ELSEVIER, Renewable and Sustainable Energy Reviews, pp.1553-1563, 2015.
[11] D. Dong, J. Herbert, "Energy Efficient VM Placement Supported by Data Analytic Service", In Proc. 13th IEEE/ACM Int. Symp. on Cluster, Cloud, and Grid Computing, pp. 648-655, 2013.
[12] R. Xie, X. Jia, K. Yang, B. Zhang, "Energy Saving Virtual Machine Allocation in Cloud Computing", IEEE, In Proc. 33rd International Conference on Distributed Computing Systems Workshops, pp. 132-137, 2013.
[13] E. Feller, C. Rohr, D. Margery, C. Morin, "Energy Management in IaaS Clouds: A Holistic Approach", IEEE, In Proc. 5th In.l Conference on Cloud Computing, 2012.
[14] X. Zhao, N. Jamali, "Energy-aware resource allocation for multicores with per-core frequency scaling", Springer, Journal of Internet Services and Applications, pp.1-15, 2014.
[15] A.M. Qawasmeh, S. Pasricha, A.A. Maciejewski, "Siegel H J. Power and Thermal-Aware Workload Allocation in Heterogeneous Data Centers", IEEE Trans. Computers, vol. 64, no.2, pp. 477-491, 2015.
[16] S. Singh, I. Chana, "Energy based Efficient Resource Scheduling: A Step Towards Green Computing", International Journal of Energy, Information and Communications, vol. 5, no.2, pp. 35-52, 2014.
[17] R. Buyya, A. Beloglazov, J. Abawajy ,"Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges", 2010.
[18] F. Farahnakian, P. Liljeber, J. Plosila, "Energy-Efficient Virtual Machines Consolidation in Cloud Data Centers using Reinforcement Learning", In Proc. 22nd Euromicro Int. Conference on Parallel, Distributed, and Network-Based Processing, pp. 500-507, 2014.
[19] K. Gupta, V. Katiyar, "Energy Aware Virtual Machine Migration Techniques for Cloud Environment", International Journal of Computer Applications, vol. 141, no.2, pp, 11-16, 2016.
[20] N. Tziritas , C.Z. Xu, T. Loukopoulos, S.U. Khan, Z. Yu, "Application-aware Workload Consolidation to Minimize both Energy Consumption and Network Load in Cloud Environments", 2012.

AUTHOR PROFILE

Er. Kamali Gupta did her B.tech from Y.M.C.A Institute, Faridabad in 2007 and received her M.Tech degree from Maharishi Markandeshwar University, Mullana in 2011. Currently, she is pursuing her Ph.D in area of Cloud Computing. She has work as System Engineer with Infosys Technologies Limited till 2010 and subsequent to it, has provided her services to Geeta Institue of Management and Technology, Kurukshetra. Currently, she is working with maharishi Markandeshwar University, Sadopur as Assistant Professor in CSE department. Er. Kamali underwent various workshops and faculty development programmes conducted by organizations like IIT Bombay, IIT Kharagpur, NITTTR Chandigarh, IBM etc. She has guided many M.Tech thesis and obtained many Microsoft Certifications in technical domains. She has many professional affiliations to her credit like IEI, IAENG etc. and has guided many students registered under her. Her research interests include cloud computing and data structures.

Dr. Vijay Katiyar is currently working as Dean Academic Affair MMU, Mullana. He did his B.Tech, M.Tech, Ph.D. in Computer Science and Engineering. He has a teaching and research experience of more than 20 years. Currently, he is serving Maharishi Markandeshwar University as Dean Academic affairs & admissions. Under his able supervision, 1 Ph.D has been awarded, 4 have been submitted more than 25 M.Tech thesis are completed. He has about 45 research publications to his credit. He has obtained IEEE membership subscription and bears a number of professional affiliations. His core interest areas includes intelligent systems and wireless sensor networks. He is associated with Punjab University, Chandigarh research journal as reviewer. More information about Dr. Vijay Katiyar can be reached by accessing the link *https://in.linkedin.com/in/vijay-kumat-katiyar-30137142.*