

Energy Efficiency Model for Cloud Computing

Arindam Banerjee, Prateek Agrawal and N. Ch. S. N. Iyengar

School of Computing Science and Engineering, VIT, India

*arindam.banerjee2012@vit.ac.in, prateek.agrawal2012@vit.ac.in,
nchsniyengar48@gmail.com*

Abstract

Cloud computing is an “evolving paradigm” that has redefined the way Information Technology based services can be offered. It has changed the model of storing and managing data for scalable, real time, internet based applications and resources satisfying end users’ needs. More and more remote host machines are built for cloud services causing more power dissipation and energy consumption. Over the decades, power consumption has become an important cost factor for computing resources. In this paper we will investigate all possible areas in a typical cloud infrastructure that are responsible for substantial amount of energy consumption and we will address the methodologies by which power utilization can be decreased without compromising Quality of Services (QoS) and overall performance. We also plan to define the scope for further extension of research from the findings we would have from this paper.

Keywords: *Cloud computing, energy efficiency, scheduling, cluster*

1. Introduction

Cloud Computing has emerged as a new consumption and virtualization model for the high cost computing infrastructures and web based IT solutions. Cloud provides suitable, on-demand service, elasticity, broad network access, resource pooling and measured service [1] in highly customizable manner with minimal management effort. The application of low-cost computing devices, high-performance network resources, huge storage capacity, semantic web technology, SOA (Service Oriented Architecture), usage of API (Application Programming Interfaces), *etc.*, have helped in the swift growth of cloud technology. A cloud infrastructure generally encapsulates all those existing technologies in a web service based model to offer business agility, improved scalability and on demand availability. The rapid deployment model, low start up investment, pay-as-you-go scheme, multi-tenant sharing of resources are all added attributes of cloud technology due to which major industries tend to virtualization for their enterprise applications [2].

Cloud applications are deployed in remote data centers (DCs) where high capacity servers and storage systems are located. A fast growth of demand for cloud based services results into establishment of enormous data centers consuming high amount of electrical power. Energy efficient model is required for complete infrastructure to reduce functional costs while maintaining vital Quality of Service (QoS). Energy optimization can be achieved by combining resources as per the current utilization, efficient virtual network topologies and thermal status of computing hardwares and nodes.

On the other hand, the primary motivation of cloud computing is related to its flexibility of resources. As more and more mobile devices are getting considered as major consumption points for remote users in mainstream business, power management has been a bottleneck for proper functioning of services at users end. A trade-off between energy consumed in computation and the same in communication has been the critical aspect to be considered for mobile clients also.

In this paper we plan to consolidate all the plausible aspects of energy efficient infrastructure model for cloud data centers while considering performance bottlenecks for the same.

2. Energy Consumption Analysis

To calculate the amount of energy consumed by data centers, two metrics were established by Green Grid, an international consortium [10]. The metrics are Power Usage Effectiveness (PUE) and Data Centre Infrastructure Efficiency (DCiE) as defined below:

$$\begin{aligned} \text{PUE} &= \text{Total Facility Power} / \text{IT Equipment Power} \\ \text{DCiE} &= 1/\text{PUE} = (\text{IT Equipment Power} / \text{Total Facility Power}) \times 100\% \end{aligned}$$

The IT equipment power is the load delivered to all computing hardware resources, while the total facility power includes other energy facilities, specifically, the energy consumed by everything that supports IT equipment load.

In cloud infrastructure, a node refers to general multicore server along with its parallel processing units, network topology, power supply unit and storage capacity. The overall energy consumption of a cloud environment can be classified as follows [9]:

$$E_{\text{Cloud}} = E_{\text{Node}} + E_{\text{Switch}} + E_{\text{Storage}} + E_{\text{Others}}$$

Consumption of energy in a cloud environment having n number of nodes and m number of switching elements can be expressed as:

$$E_{\text{Cloud}} = n (E_{\text{CPU}} + E_{\text{Memory}} + E_{\text{Disk}} + E_{\text{Mainboard}} + E_{\text{NIC}}) + m(E_{\text{Chassis}} + E_{\text{Linecards}} + E_{\text{Ports}}) + (E_{\text{NASServer}} + E_{\text{StorageController}} + E_{\text{DiskArray}}) + E_{\text{Others}}$$

3. Energy Efficiency in Cloud Infrastructures

Building an energy efficient cloud model does not indicate only energy efficient host machines. Other existing components of a complete cloud infrastructure should also be considered for energy aware applications. Several research works have been carried out to build energy efficient cloud components individually. In this section we will investigate the areas of a typical cloud setup that are responsible for considerable amount of power dissipation and we will consolidate the possible approaches to fix the issues considering energy consumption as a part of the cost functions to be applied.

3.1. Energy Efficient Hardware

One of the best approaches to reduce the power consumption at data centre and virtual machine level is usage of energy efficient hardwares at host side. International standard bodies such as: European TCO Certification [3], US Energy Star [4] are there

to rate energy efficient consumer products. The rating is essential to measure the environmental impact and carbon footprint of computer products and peripherals. New electronics materials like solid-state drives are more power efficient than common hard disk drives but that are costly. Some existing mechanisms for saving energy at hardware level are given below:

- SpeedStep® [5]: This is Intel's wireless technology to adjust CPU power dynamically based upon the performance demand. It works in five usage modes: standby mode, voice communication, multimedia, data communication, and multimedia and data communication. SpeedStep® technology can change voltage as well as frequency on-the-fly saving additional power but still provides the necessary service to run rich applications.
- PowerNow!™ [6]: This is AMD's power saving technology which can manage power consumption instantly, on-the-fly by controlling voltage and frequency independently. It can be operated in three modes namely High Performance mode, Power saver mode and automatic mode.
- Cool'n'Quiet™ [7]: AMD's Cool'n'Quiet™ technology controls the system fan, voltage and clock speed of the processor's cores based on the system temperature.
- Demand Based Switching® [8]: Intel's SpeedStep® technology's operations can be practically implemented more frequently by enabling more-granular demand-based switching and on demand power to performance balance.

All these technologies are able to sense lack of machine interaction and then different hardware parts can incrementally be hibernated or put in sleep mode to save energy.

3. 1. Energy Efficient Resource Scheduling

Several research works have been carried out on energy efficient resource scheduling in virtual machines and grid systems. Many aspects such as: power consumption, reliability, response time etc. need to be considered while designing VM schedulers.

a) Real time tasks scheduling in multiprocessor systems:

Real time scheduling of tasks in a multiprocessor host machine has been a major issue to be considered. Xian C. *et al.*, [11] presented an energy aware, real-time task scheduling method in multiprocessor system that supports Dynamic Voltage Scaling (DVS). It is a polynomial time heuristic approach for building probabilistic model of load balancing issue. Task scheduling can be of two types [13]: i) partitioned scheduling – where assignment of each individual task to a particular processor is fixed and ii) non-partitioned (dynamic) scheduling – where a long term global scheduler assigns the jobs to processors

on-the-fly from a ready queue. Several heuristics have been proposed in literature for tasks allocation having their advantages and disadvantages [12].

In a multiprocessor system, there is a set of pre-emptive, independent, real time tasks denoted as $T = \{T_1, T_2, \dots, T_n\}$ and a set of processor cores denoted as $P = \{P_1, P_2, \dots, P_m\}$ where each core has a finite range of discrete frequencies $F = \{f_1, f_2, \dots, f_k\}$. Figure 1 shows how a task T_i is shared amongst available processing units.

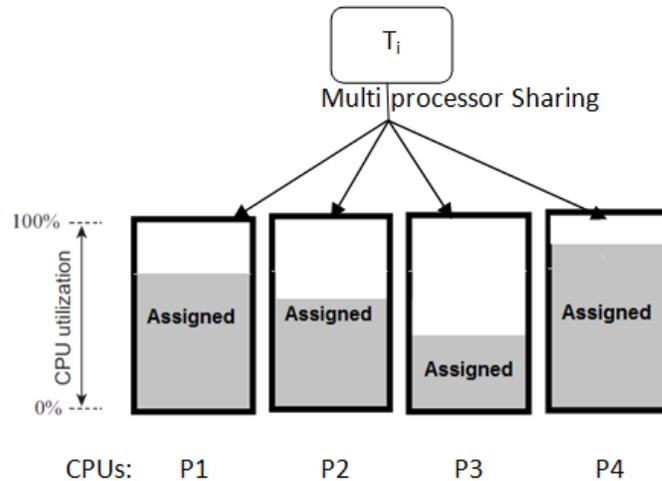


Figure 1. Multi processor sharing of a real time task

A processing unit consumes both static power (constant) and dynamic power (proportional to $c_l v^2 f$: c_l is load capacitance); dynamic power during busy period and only static power when the unit is idle. Following Early Deadline First (EDF), an optimal scheduling algorithm, DVS aims to minimize only dynamic power (as static power dissipation is constant) for continuous, unbounded frequencies and discrete, bounded frequencies of the system. In host machines, Schedulers can be customized for optimal use by utilizing known run time of virtual machines' instances, co-locating timed instances having related expiration times [14], consolidating multiple workloads considering power consumption and system performance at the same time [15]. New researches [19, 20, 21] have proposed to implement bat intelligence (BI) along with Genetic Algorithm (GA) to solve energy-aware scheduling problem with improved QoS (Quality of Service) in multiprocessing environment. Thus power consumed by scheduling operation can be reduced substantially.

b) Memory-aware Scheduling in multiprocessor systems:

Optimization of scheduling algorithm helps to reduce the energy consumption by memory and also affects the efficiency of frequency scaling. In present multicore systems, cores on the chip share resources such as caches, DRAM etc. Tasks running on one core may harmfully affect the performance of tasks on other cores and hence it may even maliciously create a Denial of Service (DoS) attack on the same chip [17]. Task assortment should be optimized by co-

scheduling them in the processor cores considering memory contention and frequency selection.

Memory aware task scheduling is based on runqueue sorting followed by frequency selection [16]. Runqueue sorting is a timeslice based multiprocessor scheduling algorithm which is a specific form of gang scheduling. For further avoidance of memory contention, frequency selection can be used which allows processor switching to a suitable frequencies for each task without causing any significant performance overhead.

c) Energy –aware Rate Monotonic Scheduling:

Dynamic Voltage Scaling (DVS) is a widely used CPU energy management approach that controls both processor's supply voltage and operating frequency. In spite of delivering significant energy gains, DVS technique increases response time while CPU speed is decreased. It causes a potential performance overhead for real time systems. For “periodic, pre-emptive, hard real-time tasks”, scheduled on multiprocessor system, DVS approach faces a real challenge for optimal energy reduction.

AlEnawy *et al.*, suggested [18] a partitioned scheduling algorithm where tasks are assigned with static, rate-monotonic priorities. They proposed an integrated approach along with rate monotonic scheduling, a partitioning heuristic, an admission control test and a speed assignment algorithm. In case of offline partitioning, task characteristics are fixed to scheduler. Here tasks can be ordered as per their periods and utilization value. Hence worst- fit allocation can be the best approach amongst other existing heuristics. In case of online partitioning, tasks arrive dynamically to the scheduler making the process complicated to handle. Here task scheduling quality of worst-fit approach degrades fast. Hybrid approach that reserves subset of processors for light tasks (tasks having small utilization factor) seems to be efficient to maintain a consistent performance in this scenario.

3. 2. Energy Efficient Clusters of Servers

Power dissipation is primarily reduced by optimal CPU utilization and tasks scheduling. However other cluster components such as memory, storage discs, network peripherals etc. also consume power and hence a VM having idle CPU may still use considerable amount of energy. Figure 2 shows a typical cloud cluster structure. New approaches aim to reduce the energy consumption as a whole at clusters of servers while considering system's latency and throughput.

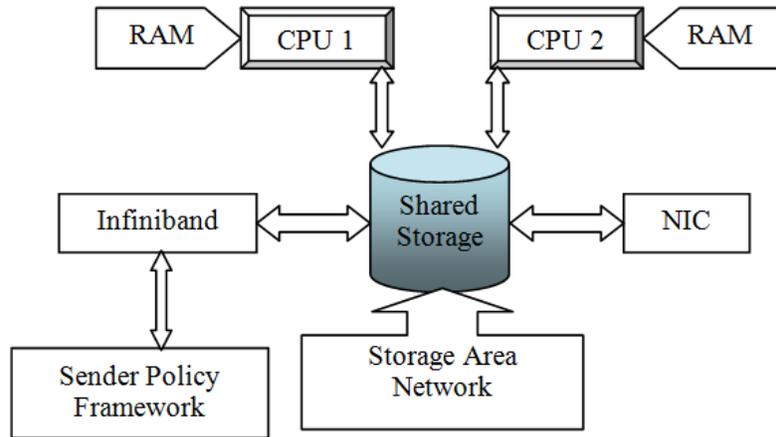


Figure 2. A Typical Cloud Cluster Structure

a) Resource management architecture:

At server side, the operating system of the machine hosts several separate instances of virtual machines. Optimal resource management architecture should be built based upon the energy estimation in different nodes of a server cluster. Depending upon external situation and workload, the cluster can be easily affected by overloading and overheating despite sufficient cooling system. While a complete shutdown of cluster causes unwanted business downtime, it is the operating system that should take care of auto-scaling of power demand from different cluster components. Dynamic Thermal Management [22] is a technique that controls power dissipation in high performance, server processing unit and provides low “worst case power consumption” with no or little impact on performance.

Cluster’s network infrastructure is a major area of power dissipation that holds a substantial share of operating cost. Balancing of QoS and resource utilization during outage can also be a descent way of energy management in clusters. Policies are developed for resource management in economic way [23] where cluster always checks for system’s work load and allocates resources by calculating their effects on system’s overall performance. A greedy allotment method can be used to estimate supply – demand tradeoffs for an efficient allocation of resources.

b) Dynamic Server Provisioning and load dispatching:

In order to save energy, Dynamic Server Provisioning method is useful in switching off unnecessary and idle hosts in a cluster [24]. As the number of internet services is increasing rapidly, the servers are also increasing in number to host those services, resulting huge amount of power dissipation in form of heat. Dynamic Server Provisioning algorithms [25] are designed to turn off extra servers and to allow the cluster to run on minimal number of host machines to satisfy the service load. Thereafter, load dispatching technique effectively distributes the current load among available servers. These techniques can be implemented in servers operating request-response type

services (example: web services) as well as host machines connected to huge number of long lived TCP connections.

For multi-tier internet services, queuing method can be implemented in dynamic provisioning technique that can be defined as a proactive and reactive approach to estimate short-term and long-term workload fluctuations in cluster. It can predict minimum capacity required for maintaining the required QoS and can also balance sudden surges in server load.

3.3. Energy efficient Network Infrastructure in cloud

Minimizing energy consumption in various elements of cloud computing such as storage and computation has already been given importance by the researchers but the issue of energy minimization in network infrastructure is not given as much importance.

Network in a cloud environment can be of two types - wireless network and wired network. According to ICT energy estimates in [26] the radio access network consumes a major part of the total energy in an infrastructure and the cost incurred on energy consumption is sometimes comparable with the total cost spent on personnel employed for network operations and maintenance. Jia *et al.*, [27] provided a thorough study on routing protocols for saving energy consumption in sensor networks and wireless ad-hoc networks. Topology Control Protocols such as Geographic Adaptive Fidelity (GAF) and Cluster Based Energy Conservation (CEC) were also presented.

Micro Sensor Architecture comprises of four components- digital processing, power supply, sensing circuitry and radio transceiver of which radio transceiver consumes maximum energy while sensing and data processing consume negligible energy. The sensor is always in one of the following states –sleeping, transmit, receive and idle. In order to achieve energy savings, sensors need to put in the sleeping state as the other three states consume considerable amount of energy. GAF and CEC protocols identify redundant nodes and turn off them to conserve energy.

a) Geographic Adaptive Fidelity Protocol:

In GAF protocol [28], equivalent nodes are found out by using their geographical information and then their radios are turned off which saves energy. However for communication between a pair of nodes, nodes which are equal may not be equal for communication between a different pair. This problem is addressed by dividing the whole network into virtual grids which have the property that all nodes in adjacent grids can communicate with each other. All nodes within a single grid are equivalent.

Also the nodes in GAF protocol always switch among one of the three states – sleeping, discovery and active as shown in Figure 3. Initially a node is in discovery state with its radio turned on and it exchanges messages with its neighbors. A node in active state and discovery state can switch to sleeping state whenever it finds an equivalent node which can perform routing.

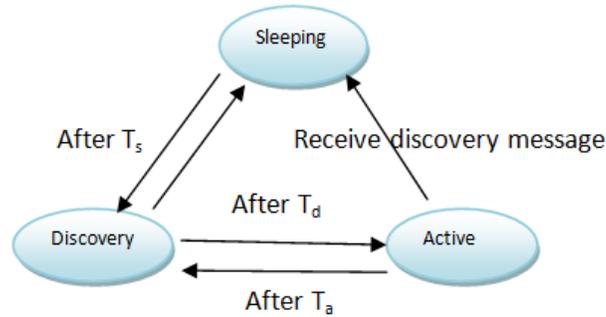


Figure 3. State transition in GAF

b) Cluster Based Energy Conservation Protocol:

One of the disadvantages with GAF protocol is that it needs global location information which may not be available every time. Also it is very conservative because it guesses its connectivity instead of directly measuring it which leads to less energy savings. Another protocol namely Cluster based Energy Conservation overcomes these disadvantages since it is independent of location information and it itself directly and adaptively measures network connectivity and finds network redundancy more accurately thus saving more energy. CEC operates in following steps-

- Determining Network Redundancy :-

It operates by organizing nodes into clusters which are overlapping. A cluster can be viewed as a circle around the node called cluster head whose radio transmission range defines the radius of the circular shaped cluster. A cluster head is defined in such a way that it reaches each node in a cluster in just one hop.

Another node called gateway node interconnects two or more clusters is a member of multiple clusters and overall network connectivity is ensured by it.

Third type of node in a cluster is called ordinary node and is the redundant one because it neither the gateway node nor the cluster head.

- Distributed Cluster formation :-

A cluster selects a cluster-head and broadcast node by broadcasting discovery messages to its neighbors. A discovery message is a combination of node ID, its cluster ID and its estimated lifetime. A cluster formation begins with the cluster head selection and then the gateway node selection.

A node having the longest lifetime selects itself as the cluster head and this information is obtained by exchanging discovery messages. The node id of the cluster head becomes the cluster id of each node in a cluster.

A primary gateway node is one that can hear multiple cluster heads and the one hearing the gateway node and cluster head is secondary gateway node. Some of the redundant gateway nodes among multiple gateway nodes between

two clusters are suppressed by CEC to conserve energy. These redundant gateway nodes are selected according to certain rules. Primary gateway nodes have higher priority and are preferred over secondary because only one primary node can connect adjacent clusters. Finally redundant nodes are powered off after selecting cluster-heads and gateway-nodes thus conserving energy [28].

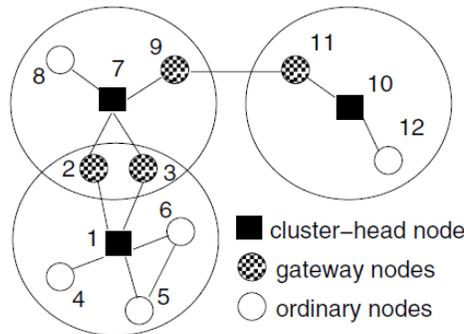


Figure 4. Cluster formation in CEC protocol

Energy savings in wired networks is not given as much importance as compared to energy savings in wireless networks. In fact, concepts applied for energy conservation in wireless networks cannot directly be applied in wired networks such as turning the nodes power off when not required to save energy because of high volumes and rate of data traffic and further nodes need to satisfy quality of constraint requirements. Also in a wired network, 60% of the energy consumed may be due to peripheral devices such as link drivers.

4. Future Scope and Research Direction

In this paper we have consolidated all three general areas of power management in cloud infrastructure and a comprehensive study has been carried out to investigate the possible scope of improvements in those areas. Future scopes for further research are there in designing more power efficient hardware, scheduling algorithm, cluster configuration as well as energy-aware network protocols.

a) Proposal based upon energy efficient scheduling:

Efficient virtual machine devices often come as proprietary hardwares from different vendors. However there are immense opportunities in designing more power efficient scheduling algorithms and cluster configuration in cloud environment. As the number of internet services and workload on host machines are increasing rapidly day by day, existing scheduling methods are becoming less effective very often. This is high time to implement new and effective methodologies in upcoming host architecture. In continuation of this paper we plan to investigate the power aware scheduling issues more minutely and will look forward for implementing hypergraph bipartitioning approach in cloud environment.

While scheduling the jobs in virtual machines, the available jobs can be represented as acyclic hyper graph $G = (V, E)$. Individual jobs are represented as nodes (set of vertices, V) of the graph where inter process dependency is represented as the graph edges (set of hyper edges, E). In hypergraph, an edge e can be connected to more than

two vertices, which means, more than two processes can be interrelated. For effective scheduling, we would focus on grouping the tasks or processes which are closely interrelated. Interrelation is measured by number of hyper edges. So, the aim of the partitioning would be achieving a minimized net cut. The bipartition of the graph aims to separate the processes (vertices) into two groups in such a way that the partitioning line cuts the minimum number of nets (edges) among processes. This bipartitioning can be iteratively applied to achieve an efficient k-way partitioning that can map the processes into one of the k disjoint partitions to reduce the cost function. A hypergraph can be represented in a matrix form as shown below:

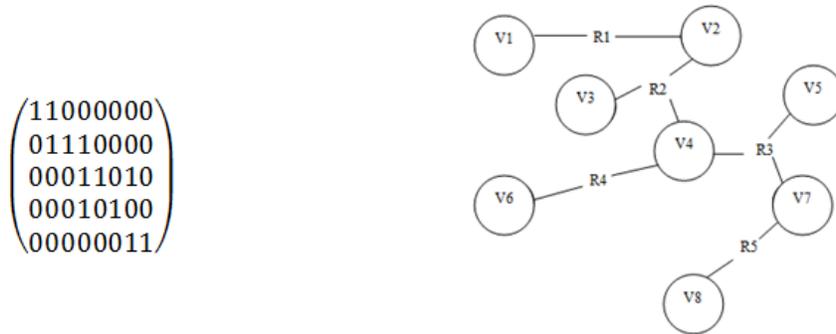


Figure 5. A hyper graph and its matrix representation are shown above. Rows of the matrix represent the hyper edges and columns represent the vertices attached to that edge

In order to make hypergraph partitioning, we would use Fiduccia-Mattheyses (FM) approach [29] – an iterative, near-linear-time heuristic based on branch and bound [30] technique. FM algorithm moves vertices into partitions based upon gain. In each pass, FM algorithm searches for the neighboring node of the given partition and calculate the best possible gain. Then one node is moved and it continues until there is little or no improvement.

By using FM algorithm iteratively, we would get k-way partitioning $P = (p_0, p_1, \dots, p_{k-1})$ where each $p_i \in P$ is a set of vertices such that, $\bigcup_{i=0}^{k-1} p_i = V$ and $\bigcap_{i=0}^{k-1} p_i = \emptyset$. Interdependent tasks (vertices) would be mapped into each partition and each partition can be allotted to a single processing unit to execute those tasks. As all partitioned tasks are nearly independent to other partitions, there would be minimum inter process communication between any two processes from two different partitions. It leads to better job scheduling considering energy efficiency in cloud environment.

b) Proposal based upon energy efficient cluster configuration:

Energy efficient cluster configuration cannot be approached as a single methodology. Cluster comprises of processing unit, memory, storage unit, network and other peripherals. Each component has its own way to address power dissipation. As the service load on remote clusters is increasing, new approaches should meet all the possible areas of energy consumption in a particular cluster. Implementing high performance hardware may increase performance but it also tends to more power dissipation. Efficient cooling mechanism, energy aware data bus, latest processing unit, *etc.*, can also help in reduction of power utilization in clusters.

5. Conclusion

In this paper we have investigated the need of power consumption and energy efficiency in cloud computing model. It has been shown that there are few major components of cloud architecture which are responsible for high amount of power dissipation in cloud. The possible ways to meet each sector for designing an energy efficiency model has also been studied. Finally we have shown the future research direction and the continuity of this work for next level implementation.

References

- [1] P. Mell and T. Grance, "The NIST definition of cloud computing", National Institute of Standards and Technology, vol. 53, no. 6, (2009).
- [2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", Tech. rep., (2009) February, UC Berkeley Reliable Adaptive Distributed Systems Laboratory.
- [3] European TCO Certification, <http://www.tcodevelopment.com>.
- [4] Energy Star, <http://www.energystar.gov>, <http://www.euenergystar.org>.
- [5] Intel whitepaper "Wireless Intel SpeedStep Power Manager: optimizing power consumption for the Intel PXA27x processor family".
- [6] "AMD PowerNow!™ Technology: dynamically manages power and performance", Informational white paper.
- [7] Computer Power User Article - White Paper: AMD Multi-Core Processors.
- [8] Intel Software Network, "Enhanced Intel SpeedStep® Technology and Demand-Based Switching on Linux", (2008), <http://softwarecommunity.intel.com/articles/eng/1611.htm>.
- [9] L. Luo, W. Wu, D. Di, F. Zhang, Y. Yan and Y. Mao, "A Resource Scheduling Algorithm of Cloud Computing based on Energy Efficient Optimization Methods", IEEE Green Computing Conference (IGCC), 2012 International, (2012) June 4-8, San Jose, CA, pp. 1 – 6.
- [10] C. Belady, "The green grid data center power efficiency metrics: PUE and DCiE", White paper: Metrics & Measurements, (2007).
- [11] X. Changjiu, L. Yung-Hsiang and L. Zhiyuan, "Energy-Aware Scheduling for Real-Time Multiprocessor Systems with Uncertain Task Execution Time", Proc. 44th Annual Conf. Design Automation, San Diego, USA, (2007) June, pp. 664–669, ACM.
- [12] J. Lopez, J. Diaz, M. Garcia and D. Garcia, "Worst-Case Utilization Bound for EDF Scheduling on Real-Time Multiprocessor Systems", Euromicro Workshop on Real-Time Systems, (2000), pp. 25-33, ACM.
- [13] S. Kato, "A Fixed-Priority Scheduling Algorithm for Multiprocessor Real-Time Systems", Chapter 8, Book: Parallel and Distributed computing, (2010) January.
- [14] T. Knauth and C. Fetzer, "Energy-aware Scheduling for Infrastructure Clouds", Proceedings of IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom), IEEE Computer Society Washington, USA, (2012), pp. 58-65.
- [15] I. Goiri, F. Julia, R. Nou, J. L. Berral, J. Guitart and J. Torres, "Energy-aware Scheduling in Virtualized Data centres", 2010 IEEE International Conference on Cluster Computing, Crete , (2010) September 20-24, pp. 58-67.
- [16] A. Merkel and F. Bellosa, "Memory-Aware Scheduling for Energy Efficiency on Multicore Processors", proceeding Workshop on Power Aware Computing and Systems (HotPower'08), San Diego, USA, (2008), pp. 123–130.
- [17] T. Moscibroda and O. Mutlu, "Memory performance attacks: denial of memory service in multi-core systems", proceedings of 16th USENIX Security Symposium on USENIX Security Symposium, (2007) August.

- [18] A. T. AlEnawy and H. Aydin, "Energy-Aware Task Allocation for Rate Monotonic Scheduling", 11th IEEE Real Time and Embedded Technology and Applications Symp. (RTAS'05), San Francisco, USA, (2005) March, pp. 213–223, IEEE.
- [19] B. Malakooti, S. Sheikh, C. Al-Najjar and H. Kim, "Multi-objective energy aware multiprocessor scheduling using bat intelligence", Journal of Intelligent Manufacturing, vol. 24, Issue 4, (2013) August, pp. 805-819, springer.
- [20] Q. -H. Nguyen, P. D. Nien, N. Hoai, N. N. Huynh and T. N. Thoai, "A Genetic Algorithm for Power-Aware Virtual Machine Allocation in Private Cloud", Information and Communication Technology, vol. 7804, Information Systems and Applications, (2013) February, pp. 170.
- [21] C. -t. Ying and Y. Jiong, "Energy-aware Genetic Algorithms for Task Scheduling in Cloud Computing", 2012 Seventh ChinaGrid Annual Conference, (2012) September 20-23, Beijing, pp. 43-48, IEEE.
- [22] D. Brooks and M. Martonosi, "Dynamic Thermal Management for High-Performance Microprocessors", proceedings of the Seventh International Symposium on High-Performance Computer Architecture (HPCA-7), (2001) January, pp. 171-182, ACM.
- [23] J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat and R. P. Doyle, "Managing Energy and Server Resources in Hosting Centres", Proc. 18th ACM Symp. Operating System Principles, Banff, Canada, (2001) October, pp. 103-116, ACM.
- [24] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao and F. Zhao, "Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services", Proc. 5th USENIX Symp. Networked Systems Design and Implementation, San Francisco, USA, (2008) April, pp. 337–350, ACM.
- [25] B. Urgaonkar, P. Shenoy, A. Chandra, P. Goyal and T. Wood, "Agile dynamic provisioning of multi-tier Internet applications", ACM Trans. Auton. Adap. Syst., (2008), pp. 31–39.
- [26] D. Sarokin, Question: energy use of Internet. <http://uclue.com/?xq=724>, (2007).
- [27] X. Jia, D. Li and D. Du, "QoS Topology Control in Ad hoc Wireless Networks", Proc. IEEE INFOCOM'04, Hong Kong, China, IEEE, (2004).
- [28] X. Ya, S. Bien, Y. Mori, J. Heidemann and D. Estrin, "Topology control protocols to conserve energy in wireless ad hoc networks", IEEE Transaction on Mobile Computing, (2003).
- [29] C. M. Fiduccia and R. M. Mattheyses, "A linear-time heuristic for improving network partitions", 19th Conference on Design Automation, IEEE, (1982).
- [30] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey", Operations research, (1966), April 14, pp. 699-719.

Authors



Arindam Banerjee (b 1987) is currently pursuing MTech in Computer Science and Engineering from VIT University, Vellore-632014, Tamil Nadu, India. He completed his graduation in Electronics and Telecommunication Engineering from Nagpur University, India. He has worked in TATA Consultancy Services, a multinational IT company, for two and half years as Systems Engineer.



Prateek Agrawal (b 1989) is currently pursuing MTech in Computer Science and Engineering from VIT University, Vellore-632014, Tamil Nadu, India. He completed his graduation in Computer Science and Engineering from RGPV University, Madhya Pradesh, India.



Dr. N. Ch. S. N. Iyengar (b 1961) is currently Senior Professor at the School of Computing Science and Engineering at VIT University, Vellore-632014, Tamil Nadu, India. His research interests include Agent-Based Distributed secure Computing, Intelligent Computing, Network Security, Secured Cloud Computing and Fluid Mechanics. He has authored several textbooks and had nearly 167 research publications in reputed peer reviewed international journals. He served as PCM/reviewer for many international and IEEE conferences. He is Editor in Chief for International Journal of Software Engineering and Applications(IJSEA) of AIRCC, Guest Editor for Special Issue on Cloud Computing and Services of *Int'l J. of Communications, Network and System Sciences* and Editorial Board member for International Journals like **IJConvC (Inderscience -China), IJCA (USA) etc.**

