Overview of Green Cloud Architecture

Sharmila S. Patil
Assis Prof.
M.Tech (CSE),
Walchand Institute of Technology, Solapur

Priyanka Pattenshetti
Student M.E (CSE)
G.H.Raisoni college of Engineering and management,Pune

ABSTRACT
Currently the Cloud computing technology is on the verge of spurring an information revolution in all regions. It offering utility-oriented IT services to users which better suited option than a traditional methods. Cloud has millions of services based on web services. Cloud is very cost effective infrastructure for this web related services. To run and maintain cloud extremely high energy is needed. This tends to increase cost and carbon emission which reduces its efficiency. This paper discusses and analyzes some of the reason which can help in green cloud architecture. This paper includes review of static architecture energy and dynamic architecture energy issues and tries to find method to solve it.

1. INTRODUCTION
Network and internet is an essential service in current world in all areas as business, education, banking and all government sectors. Demand of high speed network is confirmation of these different types of e-transactions. The concept of cloud is introduced to fulfill these web demands. Cloud contains enormous infrastructure handling large data storage, network systems and related components. Many internet service providers such as Microsoft, Google and Yahoo are operating clouds commercially. Cloud service is utility on a pay-as-you-go basis. This can help to cut down business organizations huge investment amount for capital in acquirement and maintenance of computational resources. Cloud computing provides access to scalable infrastructure and on demand services. Users can store, access, and share any amount of information in Cloud. In everyday life most of the web activity are rely on cloud like iPad which relies upon cloud to stream video, download music and books, fetch email, Google's signature products - Gmail, Google Documents and Google Earth - are delivered from the cloud and a digital library entirely hosted by servers storing most of the world's published work, all in digitized form.

Cloud computing has three layers which cover the entire computing essential to serve web services. Each of these layers provide different services to a cloud users .The lowest layer, Cloud offerings is Infrastructure-as-a-Service (IaaS) which consists of virtual machines or physical machines, storage, and clouds. Cloud infrastructures, can also be heterogeneous, integrating clouds, PCs and workstations. Moreover, the system infrastructure can also include database management systems and other storage services. The infrastructure in general is managed by an upper management layer that guarantees runtime environment customization, application isolation, accounting and quality of service. IaaS gives access to physical resources with some software configuration.

These services constitute another layer called Platform as a Service (PaaS), offering Cloud users a development platform to build their applications. Google AppEngine [1], Aneka [2], and Microsoft Azure [3] are some of the most prominent example of PaaS Clouds. In general, PaaS includes the lower layer (IaaS) as well that is bundled with the offered service. In general, pure PaaS offers only the user level middleware, which allows development and deployment of applications on any Cloud infrastructure. As noted by Appistry.com [4], the essential characteristics that identify a Platform-as-a-Service solution includes runtime framework, abstraction and cloud services.

Software as a Service allows users to access the software from any computer which is connected to the internet, SaaS providers also constitute other layers of Cloud computing and thus, maintain the customer data and configure the applications according to customer need. This reduces cost of purchasing new software and infrastructure. The customers do not have to maintain any infrastructure or install anything within their premises. They just require high speed network to get instant access to their applications.

The Cloud deployments are classified mainly into three types: Public Cloud, Private Cloud and Hybrid Cloud. Public Cloud is the most common deployment model where services are available to anyone on Internet. To support thousand of public domain users, datacenters built by public Cloud providers are quite large comprising of thousands of servers with high speed network. Some of the famous public Clouds are Amazon Web Services (AWS), Google ppEngine, and Microsoft Azure. In this deployment, Cloud services are made available to the public in a pay-as-you-go manner. A public Cloud can offer any of the three kinds of services: IaaS, PaaS, and SaaS. For instance, Amazon EC2 is a public Cloud providing infrastructure as a service, Google AppEngine is a public Cloud providing an application development platform as a service, and Salesforce.com is public Cloud providing software as a service. public Cloud offers very good solutions to the customers having small enterprise or with infrequent infrastructure usage, since these Clouds provide a very good option to handle peak loads on the local infrastructure and for an effective capacity planning[5].

1.1 Cloud Energy Usage
Clouds have various elements using enormous energy. Mainly when user accessing Cloud data using different services such as SaaS, PaaS, or IaaS over Internet, user send information through an Internet service provider's router, connecting with gateway router within a Cloud datacenter. In datacenters and at user side data goes through a local area network and are processed on virtual machines, hosting Cloud services. Each of these computing and network devices that are directly accessed and indirect devices such as cooling and electrical devices, are the major contributors to the power consumption of a Cloud. Thus thermal and energy management are the

major issues of cloud computing system due to aggregation of computing, networking, and storage hardware, the energy consumption required to transport the data from and to the user constitute. Resulting from excessive power consumption and increased component density, large amounts of heat, along with greater than ever demand for electricity, point to two major areas of concern in cloud computing operational costs and system reliability [6,7,8]

2. RELATED WORK

Environment is very sensitive topic. Researchers are putting efforts to reduce energy use and develop energy-aware models for optimizing the total energy, including operational cost, in Cloud environments. Li et al [9] propose a cost model for calculating the total cost of ownership and utilization cost in Cloud environments. They also developed suites of metrics for this calculation. However their calculation granularity is a single hardware component. Similarly, Jung et al [10] focus on power consumed by physical hosts. Their energy consumption models do not take into account the impact of specific workloads running on specific hardware. In addition, a consumer-provider Cloud cost model has been proposed by Mach and Schikuta [11]. Their energy consumption calculation is based on the number of Java Virtual Machine (JVM) instances on each server. However, it is hard to measure the actual numbers of JVM because of the dynamic nature of JVM life cycle. Moreover, Lee and Zomaya [12] propose an energy model of Cloud tasks for developing energy-conscious task consolidation algorithms to reduce energy consumption in Cloud environments. However, the energy model simply assumes the relation between CPU utilization and energy consumption is with linear increasing. Chen et al [13] propose a linear power model that presents the behavior and power consumption from individual components to a single work node. Joule meter is a power meter for VMs [14]. This makes use of software components to monitor the resource usage of VMs and then converts it to energy consumed based on the power model of each individual hardware resource. Some of the abovementioned works have made some initial efforts in benchmarking the energy and system performance. However, none of them has identified all association between energy consumption and static as well as dynamic tasks with different configurations in Cloud environments as well as system performance. In this paper, we propose a new energy consumption model and an analysis tool for Cloud environments to address these issues.

3. GREEN CLOUD COMPUTING

Cloud computing contributes for speed up Green IT which is justified by Forrester Research. Cloud based services ranging from servers, storage solutions, business applications, Software-as-a-service and Infrastructure-as-a-service – will all contribute to leveraged IT management leading to decreased e-waste illustrating a direct correlation between cloud computing and going green. A Microsoft Study encapsulated in a whitepaper provides further evidence that cloud computing helps companies move in the direction of going green. [15] The study showcases how small businesses benefit from cloud computing. Jonathan Koomey – an efficiency expert – claims that cloud computing vendors use power, infrastructure and assets much more efficiently. Cloud computing and its obvious advantages are hard to ignore: decreased costs, enhanced efficiencies, optimized data center management, better application performance, environmental friendliness, increased capacities and flexible provisioning. Cloud computing is certainly the right solution for businesses not only to help them move toward going green but also to benefit from the host of other advantages cloud computing offers. This save fossil fuels, save on energy, save paper, minimize landfill waste, reduce e-waste, and much more. [16]Ander-son et al and in [17] define fast array of wimpy nodes (FAWN) system as novel cloud architecture for low-power data intensive computing. FAWN combines low-power CPUs with small amounts of local flash storage, and balances computation and I/O capabilities in order to provide efficient parallel data access on a large-scale. FAWN has been experimentally evaluated on various workloads. The results suggest that overall, lower frequency nodes are more efficient than conventional high-performance CPUs. The FAWN architecture seems to be unfeasible to solving problems that cannot be parallelized or whose working set size cannot be further divided to be assigned into the available memory of the smaller nodes. Caulfield et al. define in [18] “Gordon” architecture, which is an example of a low-power data-centric system. By utilizing low-power processors, flash memory, and data-centric programming systems, Gordon reduces power consumption and improves performance specifically for data-centric applications. Results of the experiments presented be able to out-perform disk-based by 1.5 to 2.5 times more performance.

3.1 Architectures for energy management

Broadly two types of energy management models of energy management are static and Dynamic Architectures for energy management

- Static architecture energy management (SAM)

Optimization in computational clouds is to create the cloud systems by using low-power components and keeping the system at the acceptable level of performance. The optimization in computational clouds is to create the cloud systems by using low-power components and keeping the system at the acceptable level of performance. Mobile and handheld technology are low power options, which can applied in most of the cloud systems, CPUs may consume 35–50% of a cloud nodes total power [19], which makes them the most energy-absorbing components of the system. The other such energy expensive components are memory modules [20]. Low-power memory and CPU components of the cloud system can effectively support the energy-aware (static power) management. Green Destiny [21] and IBM Blue Gene/L [22], are the most popular examples of HPC machines, which are composed of the low power modules.

Currently, the main researches on power-aware technology are focused on CPUs and memory banks due to the large share of the consumed system power.

- Dynamic architecture energy Management (DAM).

DAM techniques include all the methods that facilitate the run-time adaptation of a cloud system according to current resource requirements and other dynamic characteristics of the cloud system states [23]. There are two main approaches to DAM technique can divide power-aware cloud according to the two dominant power-scalable components:(a) memory, and (b) processors.

a. Power-scalable memory

Memory Management Infra-Structure for Energy Reduction (Memory MISER) developed by Tolentino et al. in [20] is an efficient solution for dynamic power-scalable memory management in cloud systems. Memory MISER utilizes a modified Linux kernel and a daemon implementation of a PID controller to on-line and off-line memory scaling in the system operating mode. Memory MISER was experimentally tested on a server with 8 processors, and 32 GB of SDRAM per processor, on parallel and sequential applications. The results of the experiments presented in [20] show that Memory MISER may reach 70% reduction in memory energy consumption and 30% reduction in total system energy consumption with a performance degradation of less than 1%.The Memory MISER assures energy efficiency on cloud systems.

b. Power-scalable processors

Power-scalable processors become one of the most promising energy-efficient off-the-shelf technologies for modern cloud systems. The dynamic power of a CMOS circuit can be reduced by implementing the following voltage and frequency scaling modules: (a) dynamic voltage scaling module (DVS) , (b) dynamic frequency scaling module (DFS) and (c) dynamic voltage and frequency scaling module (DVFS) [25]. The total power of a processor (Pt) can be expressed as a sum of a dynamic power (Pd) and the static/leakage power (Ps), i.e.: Pt = Pd + Ps . (22) The dynamic power consumption of a CMOS-based processor is proportional to the percentage of active gates (A),clock frequency (f), total capacitive load (C), and voltage (V) squared [24].Pd ≈ A * f * C * V . The static power consumption is a result of energy leakage and it is calculated even if the CPU is idle [8]. DVFS techniques allow changing the voltage and frequency supply at the cloud node's to satisfy the computational requirements specified for the publications [26]. From the experiments, traditional clouds may only achieve 5–10% of the peak performance while executing scientific applications Ge et al. propose in [27] an off-line DVFS scheduling algorithm that utilizes a weighted energy-delay product to improve the cloud energy efficiency. The weighed approach of the energy-delay product is user-driven. That is, through the weighting factors the priority can be given to the energy saving or the performance, one at the expense of the other. The results of the experiments show that energy savings may reach 30% in average with less than 5% performance reduction. Off-line scheduling is effective if cloud planners have access to applications before their execution in the system. The authors also show in [27] that off-line techniques may be a good reference methodology in a comprehensive experiment analysis of run-time techniques. Huang et al. present in [28] a run-time DVFS scheduling algorithm eco, and the associate system implementation ecod. According to [28], ecod manages application performance and power consumption in real time based on an accurate measurement of CPU stall cycles due to off-chip activities." The ecod system was tested on a cloud system and the results showed a 6% reduction in performance loss and a 3% increase in energy savings.

c. Load balancing

The main aim of load balancing (LB) methodology [29, 30] is to distribute the workload across the computing cloud to achieve optimal resource utilization, minimize the response time, and avoid overload of the system. As the result some nodes in the system can be switched to the stand-by mode or just switched off. Although the energy can be saved at low power mode or inactive nodes, the overall system performance can be adversely impacted, which may be a reason of increasing the system energy utilization. Therefore LB methodologies can be characterized as a trade off between power supply and system performance. Pinheiro et al. in [30] propose a dynamical cloud operational mode controller as develop an approach to save energy in cloud systems by dynamically turning cloud nodes on and off in a way that efficiently matches load demand. The method was implemented in two types of cloud-based systems: a network server and an operating system (OS) for clouded cycle servers. The results show that the technique can save energy by taking advantage of periods of light load in cloud-based systems. Load balancing finds limited application because light loads are the exception.

4. DISCUSSION AND CONCLUSION

High operational costs and reduced cloud system reliability resulting from excessive heat are the major barriers to sustainable growth in computing power. The problem of energy efficiency in cloud computing remains challenging mainly because of the variety of applications that need to be processed on cloud systems and a continued demand for high performance. The main methods for increasing energy-efficiency in cloud computing are: (a) the SAM technique of using low power embedded CPUs coupled with flash storage and (b) the DAM technique of using software and power-scalable components to dynamically adjust cloud power consumption, especially in the form of DVFS. The potential drawback of current SAM techniques is that improving energy efficiency by using low power components has proven to be expensive. DAM techniques have shown promise for improving energy efficiency; however, designing power-aware schedulers is not trivial. Energy savings vary significantly with application, workload, and cloud system. Power consumption in static architecture energy management (SAM) and Dynamic architecture energy Management (DAM) represents a significant proportion of total power consumption for cloud storage services at medium and high usage rates. For typical networks used to deliver cloud services today, public cloud storage can consume of the order of three to four times more power than private cloud storage due to the increased energy consumption in storage, network, cooling and transport components. Public cloud storage services are more energy efficient than private and storage on local hard disk drives where comparatively less accessed and installed more architecture. However, as the number of file downloads per hour increases, the energy consumption in transport grows and storage as a service consumes more power than storage on local hard disk drives. Cloud services are more efficient than modern midrange PCs for simple office tasks.

5. REFERENCES

[1] Google App Engine. 2010. http://code.google.com/appengine/.

[2] Vecchiola, C., Chu, X. and Buyya, R. 2009. Aneka: A Software Platform for .NET-based Cloud Computing. In High Performance & Large Scale computing, Advances in Parallel Computing, ed. W. Gentzsch, L. Grandinetti and G. Joubert, IOS Press.

[3] Microsoft Azure. 2011. www.microsoft.com/windowsazure/

[4] Charrington, S. 2010, Characteristics of Platform as a Service, Cloud Pulse blog, http://Cloudpulseblog.com /2010/02/the-essential-characteristics-of-paas.

[5] Ge, R., Feng, X., Cameron, K.W.: Performance constrained distributed DVS scheduling for scientific applications on powerCloud Computaware clouds. In: Proc. of Supercomputing Conference, p. 34(2005)

[6] Kim, K.H., Buyya, R., Kim, J.: Power aware scheduling of bag-of-tasks applications with deadline constraints on

DVS-enabled clouds. In: Proc. of CCGRID, pp. 541–548 (2007)

[7] Vasic, N., Barisits, M., Salzgeber, V., Kostic, D.: Making cloud applications energy-aware. In: ACDC. Proc. of the 1st Workshop on Automated Control for Datacenters and Clouds, pp. 37–42(2009)

[8] X. Li, Y. Li, T. Liu, J. Qiu, and F. Wang, "The method and tool of cost analysis for cloud computing," in the IEEE International Conference on Cloud Computing (CLOUD 2009), Bangalore,India, 2009, pp. 93-100.

[9] G. Jung, M. A. Hiltunen, and K. R. Joshi, "Mistral: dynamically managing power, performance, and adaptation cost in cloud infrastructures," in the International Conference on Distributed Computing Systems (ICDCS 2010), Genova, Italy, 2010, pp. 62-73.

[10] W. Mach and E. Schikuta, "A consumer-provider cloud cost model considering variable cost," in the 9th IEEE International Conference on Dependable, Autonomic and Secure Computing(DASC 2011), Sydney, Australia, 2011, pp. 628-635.

[11] Y. C. Lee and A. Y. Zomaya, "Energy efficient utilization of resources in cloud computing systems," The Journal of Supercomputing, nline First, pp. 1-13, March 2010.

[12] Q. Chen, P. Grosso, K. v. d. Veldt, C. d. Laat, R. Hofman, and H.Bal, "Profiling energy consumption of VMs for green cloud computing," in the 9th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC 2011),Sydney, Australia, 2011, pp. 768-775.

[13] A. Kansal, F. Zhao, N. Kothari, and A. A. Bhattacharya, "Virtual machine power metering and provisioning," in the 1st ACM Symposium on Cloud Computing (SoCC 2010), Indianapolis

[14] download.microsoft.com/Why_and_How_Europe_Must_Reach_for_Cloud_Computing.pdf

[15] http://www.sererra.com/Go-Cloud

[16] Andersen, D.G., Franklin, J., Kaminsky,M., Phanishayee, A., Tan,L., Vasudevan, V.: FAWN: A fast array of wimpy nodes. In: Proc of the 22nd ACM Symposium on Operating Systems Principles (SOSP), Big Sky, MT (2009)

[17] Caulfield, A.M., Grupp, L.M., Swanson, S.: Gordon: using flash memory to build fast, power-efficient clusters for data-intensive applications. In: Proc. of the 14th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS ’09) (2009)

[18] Feller, E., Morin, C., Leprince, D.: State of the art of power saving in clusters and results from the EDF case study. Institut National de Recherche en Informatique et en Automatique (INRIA) (2010)

[19] Tolentino, M.E., Turner, J., Cameron, K.W.: Memory-miser: a performance-constrained runtime system for power-scalable clusters. In: Proc. of International Conference Computing Frontiers,pp. 237–246 (2007)

[20] Warren, M.S., Weigle, E.H., Feng, W.-C.: High-density computing:a 240-processor beowulf in one cubic meter. In: Proc. Of IEEE/ACM SC2002, Baltimore, Maryland, pp. 1–11 (2002)

[21] Blue Gene/LTeam: An overview of the BlueGene/L supercomputer.In: Supercomputing 2002 Technical Papers (2002)

[22] Beloglazov, A., Buyya, R., Lee, Y.C., Zomaya, A.: A taxonomy and survey of energy-efficient data centers and cloud computing systems. In: Zelkowitz,M. (ed.) Advances in Computers. Elsevier,Amsterdam (2011). ISBN 13:978-0-12-012141-0

[23] Ge, R., Feng, X., Cameron, K.W.: Performance constrained distributed DVS scheduling for scientific applications on power aware clusters. In: Proc. of Supercomputing Conference, p. 34 (2005)

[24] Chen, G., Malkowski, K., Kandemir, M., Raghavan, P.: Reducing power with performance constraints for parallel sparse applications.In: Proc. of the 19th IEEE International Parallel and Distributed Processing Symposium, p. 231a. IEEE Comput. Soc., LosAlamitos (2005)

[25] US EPA: Report to congress on server and data center energy efficiency.Technical report (2007)

[26] Ge, R., Feng, X., Cameron, K.W.: Improvement of powerperformance efficiency for high-end computing. In: Proc. of the 1st Workshop on High-Performance, Power-Aware Computing (2005), 8 pp.

[27] Huang, S., Feng,W.: Energy-efficient cluster computing via accurate workload characterization. In: Proc. of the 9th IEEE/ACM International

[28] Symposium Cluster Computing and the Grid, pp. 68–75 (2009)

[29] Pinheiro, E., Bianchini, R., Carrera, E.V., Heath, T.: Load balancing and unbalancing for power and performance in cluster-based systems. In: Proc. of Workshop on Compilers and Operating Systemsfor Low Power (2001)