International Journal of Academic Research and Development
ISSN: 2455-4197
Impact Factor: RJIF 5.22
www.academicsjournal.com
Volume 2; Issue 5; September 2017; Page No. 912-915

A study on the growth of cloud computing infrastructure

[*1] Subhendu Chatterjee, [2] Dr. RP Singh
[1] Department of UTD, Sri Satya Sai University of Technology and Medical Sciences, Sehore, Madhya Pradesh, India
[2] Department of Computer Science & Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore, Madhya Pradesh, India

Abstract
The command over cloud computing infrastructure is increasing with the growing demands of IT infrastructure during the changed business scenario of the 21st Century. The various constraints and limitations such as server capacity, storage, bandwidth and power, pose real-time challenges on datacenters. The expansion of conventional infrastructure as per the growing demand faces a real time constraint primarily to many inconvenience and inflexibilities. These complexities drive towards various issues like costing, deployment threats and various risk pertaining to the operation.
Organizations based or operations on large scale IT infrastructure have to face these challenges accompanying in near future. Henceforth, for cost effectiveness and economy solutions, there is an enormous migration towards the cloud infrastructure. Therefore, the public cloud service providers need to evolve and develop their infrastructure to meet the challenges of the increasing demand in the IT dependent society.

Keywords: cloud computing, architecture, data

Introduction
Cloud computing is not only limited to virtualization of datacenter. It was miss conceptualized because, being in the era of cloud computing, the virtualizations of datacenters were adopted to reduce the cost. Further, at various level of resource provisioning, virtualized management technology has been evolved to adapt the larger dynamic resource allocations. This further reduced costs; but also increased the datacenter flexibility and performance, ushering in a new era of optimization technology for enterprise and public clouds based upon virtualization. Cloud computing offers the possible gateways to reduce the cost and drain resources. Also creates a new organizational requirement where various teams will be responsible for networking, computation as well as storage.
Energy efficiency and low carbon strategies have attracted a lot of concern. The goal for 20% energy efficiency and carbon reduction by 2020 drove the Information Communication Technologies (ICT) sector to strategies that incorporate modern designs for a low carbon and sustainable growth. The ICT sector is part of the 2020 goal and participates in three different ways. In the direct way, ICT are called to reduce their own energy demands (green networks, green IT), in the indirect way ICT are used for carbon displacements and in the systematic way ICT collaborate with other sectors of the economy to provide energy efficiency (smart-grids, smart buildings, intelligent transportations systems, etc.). ICT and in particular datacenters have a strong impact to the global CO_2 emissions. Moreover, an important part of the operational expenditure is due to the electricity demands.
The demand for high speed data transfer and storage capacity together with the increasingly growth of broadband subscribers and services will drive the green technologies to be of vital importance for the telecommunication industry, in the near future. Already, recent research and technological papers show that energy efficiency is an important issue for the future networks. A review of energy efficient technologies for wireless and wired networks is presented by Brown *et al* (2007). The design of energy efficient WDM (*Wavelength Division Multiplexing*) ring networks is highlighted. It is shown that energy efficiency can be achieved by increasing the capital expenditure of the network, by reducing the complexity and by utilizing management schemes. The case of thin client solutions is investigated and it is shown that employing power states in the operation of a datacenter can yield energy efficiency. Efforts have been cited related to agreeing and enabling standard efficiency metric, real time measurement systems, modeling energy efficiency, suggesting optimal designs, incorporating renewable energy sources in the datacenter and developing sophisticated algorithms for designing and managing the datacenters. These approaches have been published by various companies, experts in the field and organizations. Although the IT industry has begun "greening" major corporate datacenters, most of the cyber infrastructure on a university campus or SMEs of suboptimal energy environment and ad hoc involves a complex network, in small departmental facilities placed with clusters.

Objectives of the Research
The focus of this research work is to analyze the various power issues on the core cloud computing infrastructure along with network and storage model with provisioning parameters for optimization of resource allocation by the cloud. The main objective of this research work is discussed as below:

1. To conduct a survey about the various energy issues of the large scale cloud computing architecture.
2. Conduct the trade-off analysis using the parameters for estimation of *Service Level Agreement* (SLA) violations; cloud estimation in heterogeneous *Dynamic Voltage Frequency Scaling* (DVFS) enabled data-center for better visualization of the proposed analysis.

Review of Literature

Fan *et al* (2007) [1] aggregated power consumption of large collections of servers for different classes of applications over history data. Meisner *et al* (2009) [2] incorporated suspending and waking transitions to the power model. Lang *et al* (2010) proposed a mathematical model for the energy consumption of a Map Reduce cluster (Kavulya *et al* 2010), which adopted the workload characteristics and hardware characteristics as abstract meta-models. Poess *et al* (2008) [3] developed a TPC-C with the eagerly available data for power consumption pattern and it has been a confession of report benchmarks. Moreno *et al* (2010) [4] have addressed the importance of energy savings without degrading the performance in cloud computing, since more than a technological advance it represented a business model where the satisfaction of customers has high priority. The state of art in energy-aware computing for cloud environments shows that the initials efforts for saving energy have started primarily focused in the reduction of energy waste generated by idle servers mainly supported by VM consolidation and live migration. These, in conjunction with scheduling algorithms have boosted up two main trends: "dynamic server´s pool resizing" and "dynamic processor scaling".

Galloway *et al* (2011) [5] introduces a load balancing algorithm that balances resources across available compute nodes in a cloud with power savings in mind is introduced. Since the cloud architecture implemented by local organizations tends to be heterogeneous, this is taken into account for this proposed design.

Buysse *et al* (2011) [6] investigated that the IT infrastructure and optical network is integration of an operation facilitating the energy efficient. The proposed energy efficient routing algorithm at context level for provisioning of IT services. The IT resources are executed with the suitable originates from specific source sites e.g. datacenters. The routing approach followed is unicast, the IT service is delivery of results that are required then finding the exact location of the job execution has been chosen freely. In this scenario, IT and network resources are required to support the services, when the energy efficiency is achieved, the least energy consumption can be identified and turning off of any unused IT resources and networks.

Wang *et al* (2012) [7] have proposed a new energy-efficient multi-job scheduling model based on the Google's vast data processing framework, Map Reduce, and create the corresponding algorithm. Meanwhile, proposed individual decoding and encoding effective method and construct the individual fitness value of the servers and overall function of the energy efficiency. Also, a local search operator is introduced for searching ability of the proposed algorithm to check if the model is in order to accelerate the convergent speed and enhance.

Kim *et al* (2011) have proposed two types of approach that are illustrated as (i) a real-time service as a real-time virtual machine request to model; and (ii) virtual machines in Cloud datacenters using to provision of Dynamic Voltage Frequency Scaling (DVFS) schemes. It also proposed a various schemes of power aware profitable provisioning of soft real time services and to reduce power consumption by real time.

Kim *et al* (2009) studied the problem of real-time Cloud service framework where each real-time service request is modeled as RT-VM in resource brokers. And it investigated power-aware provisioning of virtual machines for real-time Cloud services. Simulation results shows that datacenters can reduce power consumption and increase their profit using DVS schemes. The proposed scheme, Advance-DVS and Advanced- DVS shows more profit with less power consumption regardless of system load.

Quan *et al* (2011) proposed a method that potentially reduces the energy consumption of the internal IaaS data centre. To save energy, the resources allocation by the workload consolidation and frequency adjustment is rearranged. In the reallocation algorithm, the advantage of the fact that new generation computer components have higher performance and consume less energy than the old generation is taken.

Kim *et al* (2010) evaluated Apache Hadoop on low power machines and study of the feasibility. Also proposed Augmentation and Substitution which is energy saving method to reduce energy consumption by introducing low power machines. The proposed system implements An Swer in Hadoop and experimentally studied An Swer in depth to measure the impact on performance and power savings. Furthermore, the other benchmark tools are used to study the behavior of data processing frameworks in various ways.

Research Methodology

The consumption of the datacenters resources is low average due to reason of the power efficiency. The utilized servers cannot accumulate new service applications, therefore to keep the desired Quality of Service (QoS), all the fluctuating workload needs to be acknowledged that has lead to the concert of degradation. Conversely, servers in a non-virtualized datacenter are unlikely to be completely idle, because of background tasks e.g. incremental backups or distributed databases or file-systems. An additional problem of high power consumption due to increasing density of server's components i.e. 1U, blade servers, is the heat dissipation. DVFS minimizes the number of instructions that a processor can issue in a given amount of time, and thereby it minimizes the performance too. This, in turn, increases run time for program segments which are CPU-bound.

One of the significant goals of the proposed study is to analyze the root source of energy consumption on various cloud based application and then introduce a model that will be responsible for reducing the energy consumption as well as minimize the performance loss in cloud computing system.

The first goal is accomplished by incorporating the novel design by performing the semantic evaluation of energy consumption that is primarily focused on monitoring Service Level Agreement (SLA) violation. The process is accomplished by using DVFS in the schema for energy saving approach with dual benefits e.g. i) Resource Throttling: it can

scrutinize resource utilization, memory, and wait time on both peak and off hours ii) Dynamic Component Deactivation: This technique will allow to deactivate the cloud components when in idle mode for leveraging the workload variability. However, the performance loss is scaled using network model that uses broker design and cloudlet ID mainly. And By incorporating the newly established design of energy optimization using DVFS aimed for massive task execution.

Rivoire *et al* (2007) has proposed an approach that uses *Massive Arrays of Inexpensive Disks* (MAID). They proposed the use of a small number of cache disks in addition to the MAID (*Massive Array of Idle Disks*) disks. The data in these cache disks is updated to reflect the workload that is currently being accessed. The MAID disks can then be powered down, and need only be spun up when a cache miss occurs, upon which their contents are copied onto the cache disks. This approach has several memories of the weaknesses that catches suffer on a large scale. If the insufficient cache disks are to store the entire working set of the current workload, then 'thrashing' results, with considerable latency penalties. Further, the cache disks represent a significant added cost in themselves.

There are two groups of cloud, DCTCP (*Datacenter Transfer Control Protocol*) and wide area TCP (*Transfer Control Protocol*). If it compared while using, the DCTCP delivers 90% less buffer space and better throughput than TCP, but TCP provides low latency for short flows and high burst tolerance (Alizadeh *et al* 2010).

Qian *et al* (2012) have elucidated the formulation of hardware cost to minimize energy consumption as well as cloud servers under three different models (heterogeneous, mixed hetero-homogeneous, homogeneous clusters) by considering the dynamic demand temporal. The study shows that the homogeneous model takes four times less computation time than the heterogeneous model. The energetic aggregation scheme results in 8% to 40% savings over the static aggregation scheme when the degree of aggregation is high.

Yamini *et al* (2012) proposed an approach that the cloud is considered for both private and public. It studied the energy consumption in both cloud computing. Cloud computing is using the computing power with green algorithm can enable more energy-efficient.

Significance of the Study

Low utilization of server is the biggest factor in a datacenter with low power. For example, the regular utilization of server in a Google datacenter was reported to be 30% energy efficiency (Pinheiro *et al* 2007). This fact has motivated the design of energy-proportional servers to minimize the overall power consumption (Ryckbosch *et al* 2011). State-of-the-art commercial servers are, however, not proportional to energy utilization. It is thus prudent from an energy efficiency viewpoint to have as few servers as possible turned being highly utilized with each active server. Hence, there is a strong justification for server consolidation in current datacenters. Operational cost and admission control policy in the cloud computing system are affected by its power control and VM management policies. Power management techniques control the average and/or peak power dissipation in datacenters in a distributed or centralized manner (Raghavendra *et al* 2007; Srikantaiah *et al* 2008). VM management techniques control the VM placement in physical servers as well as VM migration from a server to another one (Tang *et al* 2007; Kimbrel *et al* 2005; Verma *et al* 2008; Beloglazov and Buyya 2010). The proposed study focused on SLA-based VM management to minimize the operational cost in a cloud computing system.

The IT infrastructure provided by the datacenter owners/operators must meet various SLAs established with the clients. The SLAs may be resource related e.g., amount of computing power, space, memory/storage network bandwidth, performance related e.g., service time or throughput), or even quality of service related e.g., 24-7 availability, data security, percentage of dropped requests. Infrastructure providers often end up over provisioning their resources in order to meet the SLAs client's. Such over provisioning may increase the operational cost of the datacenters in terms of their monthly carbon emission and electrical energy bill. Therefore, in order to minimize the impact of datacenters on the environment and optimal provisioning of the resources to reduce the crucial cost incurred on the datacenter is prioritized.

Discussion

The research proposal probes various aspects of modeling the power consumption in datacenters. The proposed system attempts to investigate a complex service analysis for cloud computing which is structured as virtual machine for various resource brokers. The research also finds various power-aware policies of virtual machines for Cloud services. Simulation environment is created in java platform which shows that datacenters can predominantly reduce energy consumption and increase their profit using proposed DVFS schemes which shows maximized profit with minimized power consumption irrespective of system load. To reduce the number of processor instructions of dynamic frequency scaling can issue in given time while reducing performance. Dynamic frequency scaling has been rarely advisable as a way to conserve the power switching itself. The most power requires dynamic voltage scaling in saving, because of the fact that modern CPU and V2 components are strongly optimized for low idle states of power. It is more efficient to run in most constant voltage cases to briefly at peak speed and stay longer in a deep idle state, then it is reduced at a reduced clock rate for long time and only stay briefly in a idle state. Thus, reducing the voltage along with clock rate can change the tradeoffs.

Conclusion

The proposed study was evaluated in highly controlled research environment as performing sophisticated experiments on real time giant datacenters is almost near to impossible. However, the effect of the traffic, protocols used, considerations of downtime formulation, service level agreement, as well as provisioning schema was entirely simulated using available cloud simulators. But, simulating the considered research variable (energy) in cloud simulator and extracting the throughput from it doesn't give the actual visualization of what exactly happens in real time traffic in cloud environment. However, the reliability of the mathematical evaluation considering the effectiveness of

DVFS system is well assured and promising, but it is not experimented over wide real time test cases of dynamic internet environment considering cloud services. Currently, the proposed study is only limited to Software as a service consideration, whereas Infrastructure as a service and Platform as a Service are yet to be evaluated, which may be the scope of future direction. The stress was less on service domain as the proposed study is purely concentrated on the energy preservation schema. Extending the current study over various set of services as well as extensive testing over wide applications over datacenter can exponentially increase the reliability of the proposed study.

References

1. Fan R, Rao AR. Data Reduction techniques, Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi, 2007.
2. Meisner SP, Mani S. The State of High Performance Computing in the Cloud, Journal of Emerging Trends in Computing and Information Sciences. 2009; 3(2):262-266. ISSN 2079-8407.
3. Poess H, Aldabbas H, Alwada'n T. Comparison between cloud and grid computing: review paper, International Journal on Cloud Computing: Services and Architecture. 2008; 2(4).
4. Moreno G, Marco Conti M, Francesco DM, Andrea Passarella A. Energy conservation in wireless sensor networks: A survey, Ad Hoc Network. 2010; 7(3):537-568.
5. Galloway A, Czajkowski K, Dan A, Keahey K, Ludwig H, Pruyne J, *et al*. Web services agreement specification (WSAgreement), Global Grid Forum, 2011.
6. Buysse J, Water land A, Silva DD, Uhlig V, Rosenburg B, Hensbergen EV, *et al*. Providing a cloud network infrastructure on a supercomputer, In Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, ACM. 2011, 385-394.
7. Wang R, Stumm M, Wisniewski RW. Online performance analysis by statistical sampling of microprocessor performance counters, In Proceedings of the 19th annual international conference on Supercomputing, ACM. 2012, 101-110.