Association for Information Systems
AIS Electronic Library (AISeL)

CONF-IRM 2015 Proceedings

International Conference on Information Resources Management (CONF-IRM)

5-2015

A Policy-Based Network Management Approach for Greening the Cloud Infrastructure

Fadi Alhaddadin
Auckland University of Technology, fadi.alhaddadin@aut.ac.nz

William Liu
Auckland University of Technology, william.liu@aut.ac.nz

Jairo Gutierrez
Auckland University of Technology, jairo.gutierrez@aut.ac.nz

P29. A Policy-Based Network Management Approach for Greening the Cloud Infrastructure

Fadi Alhaddadin
Auckland University of Technology
fadi.alhaddadin@aut.ac.nz

William Liu
Auckland University of Technology
william.liu@aut.ac.nz

Jairo Gutierrez
Auckland University of Technology
jairo.gutierrez@aut.ac.nz

Abstract

Cloud computing technology is gaining great popularity as a technological enabler for different types of organisations. It offers a great deal of benefits with respect to its pay-as-you-go elasticity and capability to scale. Due to the increasing demand for cloud-based solutions, many challenges have appeared and some of them are constraints to what the technology can potentially offer. Energy efficiency is considered one of the main challenges for cloud computing technologies. This research aims to investigate the power consumption issues in current cloud computing implementations. It also proposes a new concept to find out potential trade-off solutions between energy efficiency and performance using a policy-based network management approach.

Keywords

Power Efficiency, Policy-Based Network Management, Energy Efficient Cloud Computing

1. Introduction

The term "cloud computing" refers to the provision of a computing-as-a-service model in which computing resources are made available as a utility service (Mell and Grance, 2011). It is an illusion of availability of resources as demanded by the user; resources such as CPU, memory, and I/O (Mishra et al. 2012). Cloud Computing can be classified as a new paradigm for the dynamic provisioning of computing services supported by state-of-the-art data centers that usually employ virtual machine technologies for consolidation and environment isolation purposes (Barham et al. 2003). With the continuous development and advancement in the information technology field, cloud computing has become one of the most promising technologies in recent years due to the great opportunities that it can potentially offer. Cloud computing today provides infrastructure, platform, and software as services that are made available to consumers in a pay-as-you-go model. Cloud computing is a technology which builds on decades of research in virtualization, distributed computing, utility computing, and more recently networking, web and software services (Vouk, 2008). The significant value that cloud technology offers to enterprises is twofold. Firstly, the relief from the necessity in setting up basic hardware and software infrastructure, allowing organizations to concentrate on innovation and creating business value for their services and/or products. Secondly, it cuts down the cost of the computing infrastructure as developers with innovative ideas for new internet services, using cloud technology, no longer require large capital outlays in hardware to deploy their services or labor expenses to operate them (Armbrust et al. 2010).

In order to obtain the best of cloud computing technology, it is essential that cloud service providers ensure flexibility in their service delivery to meet various consumer requirements, while keeping the customer isolated from the underlying infrastructure. However, until recently, the performance of the cloud has been the major concern in data center deployment and the fulfillment of such demand has been accomplished without much consideration to the energy consumption issue. A typical data center, consisting of 1000 racks, requires 10 Megawatts of power to operate (Rivoire et al. 2007). These levels of power consumption are significant and bound to grow as the adoption of cloud computing increases. Jonathan Koomey, in his PhD research, found that global datacenters in 2010 were responsible for 1.1% to 1.5% of the total power consumed in the world (Koomey, 2011). Greenpeace estimates that cloud computer sites could consume up to 622.6 billion KWh (Kilowatts per hour) of power (Greenpeace, 2010). On the other hand, Gartner in 2007 estimated that the Information and Communication Technologies (ICT) industry generates approximately 2% of the total global CO^2 emissions which is equal to what the aviation industry consumes (Rivoire et al. 2007). Thus, electricity consumption and carbon emission by cloud infrastructures have become a key environmental concern that needs to be addressed. This research aims to achieve the necessary balance between energy efficient cloud-based solutions and their performance using the Policy-Based Network Management (PBNM) approach. The main objective of the research is to express, deploy, and enforce power consumption policies into the cloud, and to develop a system that is able to handle policy conflicts within the context of policy-based network management.

The remaining of the paper is organized as follows: section 2 presents related work. Section 3 demonstrates the proposed approach toward overcoming the issue of power consumption in cloud computing. Section 4 discusses the simulation study and the results achieved using the proposed approach and algorithms. Finally, section 5 concludes the paper.

2. Related work

Energy consumption is considered one of the most challenging issues in the field of cloud computing. The increasing demand on cloud computing infrastructures and services has become a major environmental issue due to the amount of power it requires (Kuribayashi, 2012). Some research has been conducted into the use of cloud computing for three different services such as storage services, software services, and processing services in public and private systems. The results showed that cloud computing is not always the greenest option in terms of power consumption (Zyga, 2010).

Recently (in 2012) research conducted by Shin-ichi Kuribayashi on cloud computing has identified a need for collaboration among servers, communication networks, and power networks in order to reduce the total power consumption by the entire ICT equipment in a cloud computing environment (Kuribayashi, 2012). In order to achieve a stable collaboration among servers, communication networks, and power networks, it is important to draw and employ a steady and satisfactory management scheme which assures the required collaboration within the network.

In 2005, Agrawal, Lee, and Lobo from the IBM T. J. Watson Research Centre conducted a study which provided an overview of how the Policy Management for Automation Computing (PMAC) platform works and manages a networked system. They demonstrated the concept of a policy-driven management model used on networked systems. The outcome of the study revealed the ability of policy-based network management in reducing the burden on the human administrator in such networks by providing systematic means to create, modify, distribute, and enforce policies for managed resources (Agrawal et al, 2005). The building of a robust policy-based management framework seemed essential to achieve the PMAC goals. In 2002, Verma conducted research on the Policy-Based Management System and its usability for network administration purposes (Verma 2002). Verma's study aimed to demonstrate how network administration can be simplified using a policy-based management system. It also intended to express the main framework-related issues encountered and those that need to be considered when developing policy-based management systems such as the policy conflict issue. The policy conflict issue occurs when two or more policies are due to be enforced at simultaneously due to satisfactory conditions for both of them at the same time.

Policy-based network management principles have also been used to promote new ways of pricing network services (Gutierrez, 2008; Beltran, Gutierrez and Melus, 2010; Ray and Gutierrez, 2014). Policies allowed network providers to offer dynamic pricing based on the network context and customers had the advantage of using services without having a contract with the ad-hoc provider.

3. Proposed approach

In response to the issue of power consumption in cloud computing environments, this paper presents an algorithm that employs the consolidation approach described in (Mishra et al. 2012) along with the policy-based network management approach using dynamic provisioning and virtual machine migration techniques. The main goal of the consolidation, in this approach, is to reduce the total number of servers and/or locations that a particular data center requires at a particular time. The advent of virtualization allows consolidation of multiple applications into virtual containers hosted on a single or multiple physical servers (Padala et al. 2007). Server consolidation contributes towards avoiding the problem of server sprawl. Server sprawl refers to the situation when under-utilized servers use more resources that can be justified by their workload (Spellmann,2003). In cloud computing, the key concept that enables consolidation is the mechanism of virtual machine migration. Virtual machines can be migrated from lightly loaded hosts to other capable ones in a datacenter, which enables the freed up hosts to be switched off to save power and decrease the power consumed within the datacenter. More details on virtual machine migration can be found in (Clark et al. 2005). Figure (1) below illustrates the consolidation to three servers within a data center when two servers out of three have been switched off.

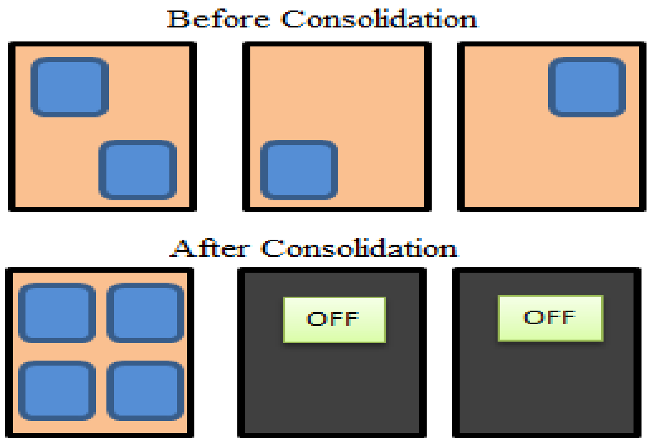

Figure 1: Consolidation Concept

However, applying the consolidation approach using virtual machine migration can pose issues and challenges in terms of quality of service (QoS). Maintaining the quality of service is one of the major challenges encountered when deciding to minimize energy consumption in any cloud-based system. In the consolidation approach, achieving the required balance between the energy consumption and the quality of service requires an accurate virtual migration process management. This management involves two major stages: the selection of the virtual machine that has to migrate and the allocation process of it on the destination host. The following section presents a description of the simulation trials and results obtained towards improving the trade-offs between power consumption and quality of service. As mentioned earlier in the paper, achieving efficiency in power consumption in a cloud-based system using the consolidation approach requires employing accurate optimization and management techniques for virtual machine migration. To that end, we propose two optimization algorithms to be simulated and compared further in our simulation stage to find out which algorithm produces a better trade-off between power efficiency and QoS.

3.1 Minimum Migration Time (MMT)

MMT (Minimum Migration Time) is a virtual machine selection policy. The MMT policy works by listing all virtual machines and sorting them according to the time they require to complete their migration. The virtual machine requiring the least time to migrate gets selected for migration. The time consumed in the migration process equals the amount of memory of the Virtual machine divided by the bandwidth of the host (Beloglazov and Buyya, 2011). The pseudo-code for the MMT policy is presented in Figure 2 below.

```
double minMetric = Double.MAX_VALUE;
for (Vm vm : migratableVms)
        {
        if (vm.isInMigration())
        {
        continue;
        }
        double metric = vm.getRam();
        if (metric < minMetric)
        {
                minMetric = metric;
                vmToMigrate = vm;
        }

        return vmToMigrate;
```

Figure 2: MMT Policy

The first step a MMT algorithm takes in its operation is listing virtual machines which are eligible for migration. Once the list of virtual machines is prepared, it looks for all virtual machines that are eligible for migration. Once all virtual machines are listed for migration, it looks for the virtual machine which requires the least time to complete their migration process and selects it for migration to another host. The process of finding the virtual machine that requires the least time to complete their migration process happens by comparing the time a virtual machine requires for migration against a certain value referred to in the pseudo-code as (MAX_VALUE).

3.2 Minimum Utilization VM selection (MU)

MU (Minimum Utilization) is another selection policy. The MU policy works by listing down all virtual machines and categorizing them according to their CPU utilization. The algorithm uses a certain value referred to as MAX_VALUE in the pseudo-code to compare the CPU utilization of virtual machine against it. The virtual machine with the least utilized CPU is the most eligible for migration. The pseudo-code for the MU policy is presented in Figure 3.

```
public Vm getVmToMigrate(PowerHost host) {
        List<PowerVm> migratableVms =
getMigratableVms(host);
        if (migratableVms.isEmpty())
        {
        return null;
        }
        Vm vmToMigrate = null;
        double minMetric = Double.MAX_VALUE;
        for (Vm vm : migratableVms) {
                if (vm.isInMigration())
                {
                continue;
                }
double metric = vm.getTotalUtilizationOfCpuMips
(CloudSim.clock())/ vm.getMips();
        if (metric < minMetric)
        {
        minMetric = metric;
        vmToMigrate = vm;
        }
        return vmToMigrate;
```

Figure 3: MU Policy

3.3 Static Threshold Value

In order to achieve the best trade-off between power efficiency and QoS that each of MMT and MU policies can offer, we find it useful to statically set the upper threshold value for algorithms to behave according to it. The Static Threshold Value (STV) plays a vital role in our approach as it helps finding the best adjustment for each of the policies towards the desired trade-off between power efficiency and QoS. The following section presents a description of our simulation trials and the results obtained towards improving the trade-off between power consumption and quality of service.

4. Simulation Studies

To test the efficiency of the proposed approach in terms of power consumption and quality of service, our simulation involved running two types of algorithms in a similar network scenario: power-aware and non-power aware algorithms. In the first run of our simulation, a non-power aware policy (NPA) was run in which no virtual machine migration facility was enabled hence no consolidation took place and all physical nodes run on their maximum power capacity throughout the simulation time. The second run of our simulation involved running two different power-aware algorithms for the goal of consolidation: Minimum Migration Time algorithm (MMT) and Minimum Utilization algorithm (MU). Each algorithm was run along with STV in order to find out the efficiency of each algorithm at different values of the upper threshold. The goal of running two different algorithms is to find out which algorithm can produce better trade-off between power consumption minimization and fewer quality of service violations when consolidation is aimed for. Discovering the more power efficient algorithm required heterogeneous experimental trials. Results of each simulation trial were monitored and recorded.

4.1 Simulation Settings

The CloudSim simulator was the tool employed for this research. Information about CloudSim simulator is available in (Rodrigo et al. 2011). The CloudSim simulator was selected for this research due to the great amount of cloud system scenarios that it offers and the flexibility it provides for modifying and tuning its current algorithms. It is a Java-based software and modifying its algorithms and scenarios requires applying basic changes to the Java code. This has made it easier to focus more on forming different combinations of "virtual machine selection" and "virtual machine allocation" algorithms and tune them towards the best possible combination for balancing the power consumption and the quality of service.

Our simulation was conducted on a simulated datacenter comprising of 50 hosts and 50 virtual machines running on them. Each host in the datacenter is designed to have one CPU core with the performance equivalent to 1860 and 2660 MIPS, 3 GB of RAM, 1 GB of storage, and 1GB of bandwidth. 4 types of virtual machines were involved in the simulation. The size of each virtual machine is 2.5 GB. Each virtual machine requires one CPU core with 2500, 2000, 1000 or 500MIPS, 870, 1740, 1740, 613 of RAM. Each virtual machine runs applications with variable workloads and they are modeled to generate CPU utilization according to uniformly distributed random workload variables provided by the CloudSim package. The initial distribution of virtual machines on hosts within the datacenter was configured according to the requested characteristics assuming full CPU utilization. Each experiment was run for 86400 seconds (24 hours). The simulation for the selected algorithms was run 9 times with 9 different tunes (Static Threshold values) to find out the best adjustments of these algorithms towards the best balance between energy consumption and quality of service.

4.2 Simulation Results and Analysis

For our benchmark experimental results, in the first simulation run when the non-power aware simulation policy was applied, no power-aware optimization policies were employed hence all hosts in the datacenter were run at the full capacity of their CPUs consuming the maximum amount of power throughout the simulation time. Therefore, the power consumed was very high and reached up to 150 Kwh and remained at the same level throughout the simulation time. However, the quality of service obtained from the NPA policy was very high as no violation to the Service Level Agreement (SLA) occurred. The reason behind the total absence of SLA violations when using the NPA policy can be explained by the lack of hosts' shutdowns in the datacenter as there was no need for VM migrations. Figure 4 presents the results obtained when the NPA policy was run.

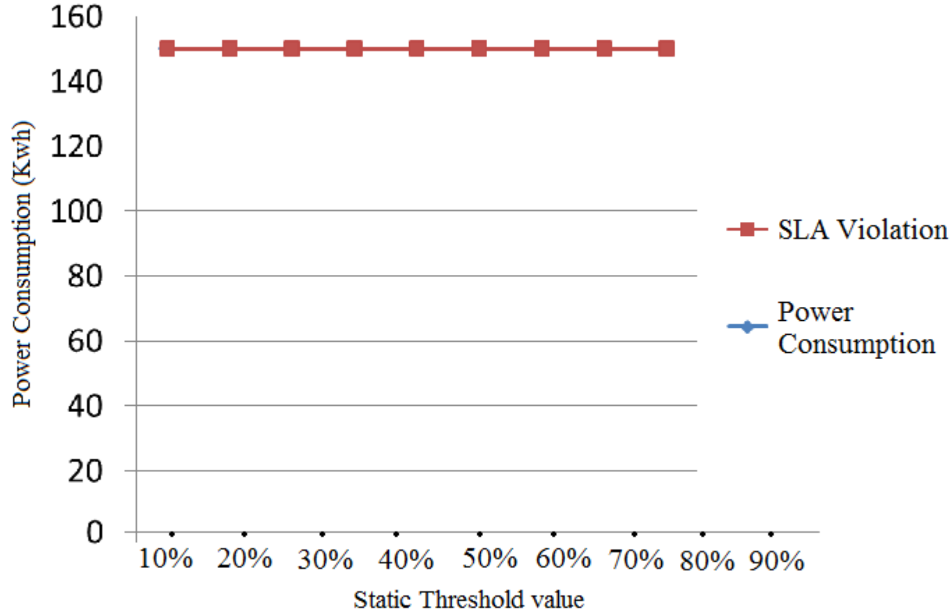

Figure 4: NPA simulation results

In the second run, when two optimization algorithms MMT and MU were employed in two separate simulations, a great improvement was noticed in terms of power efficiency due to the consolidation processes for both algorithms. Figure 5 shows the power consumption for both algorithms in 9 runs with 9 different static threshold values.

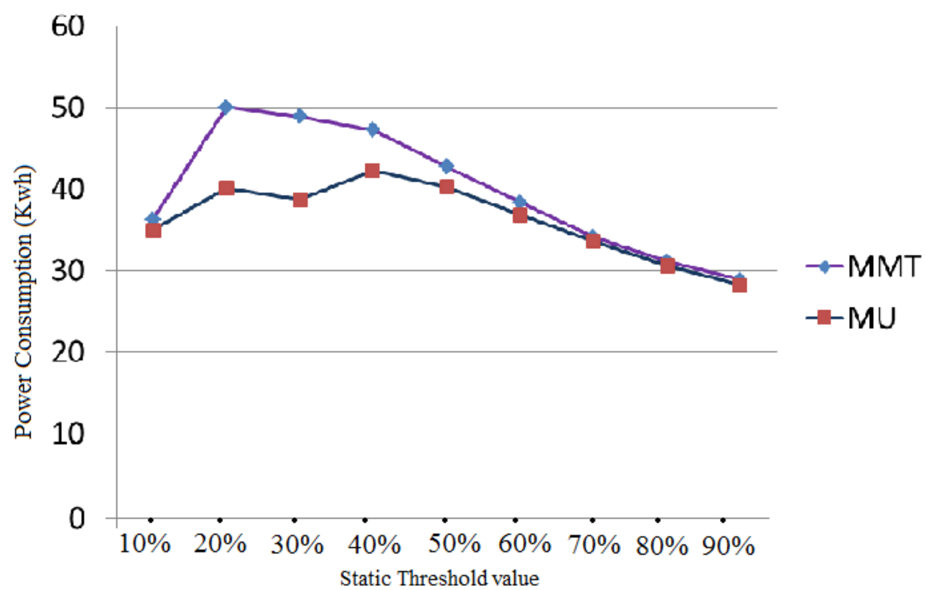

Figure 5: Power Consumption results (MMT vs. MU)

From Figure 5, it is easy to notice that the MU algorithm brings more power sufficiency than the MMT algorithm. This indicates that more consolidation processes happen due to MU than in the

MMT case. Also, the strong relationship between the Static Threshold values and the amount of power consumed is noticeable. In both algorithms, the amount of power consumed decreases with the increase in the static threshold value. However, it is also important to find out what is the effect of the consolidation processes on the QoS provided by both algorithms in order to compare the total sufficiency of both algorithms in terms of the trade-off between power efficiency and quality of service. Figure 6 presents the results of Service Level Agreement violations due to the consolidation processes in both algorithms MMT and MU.

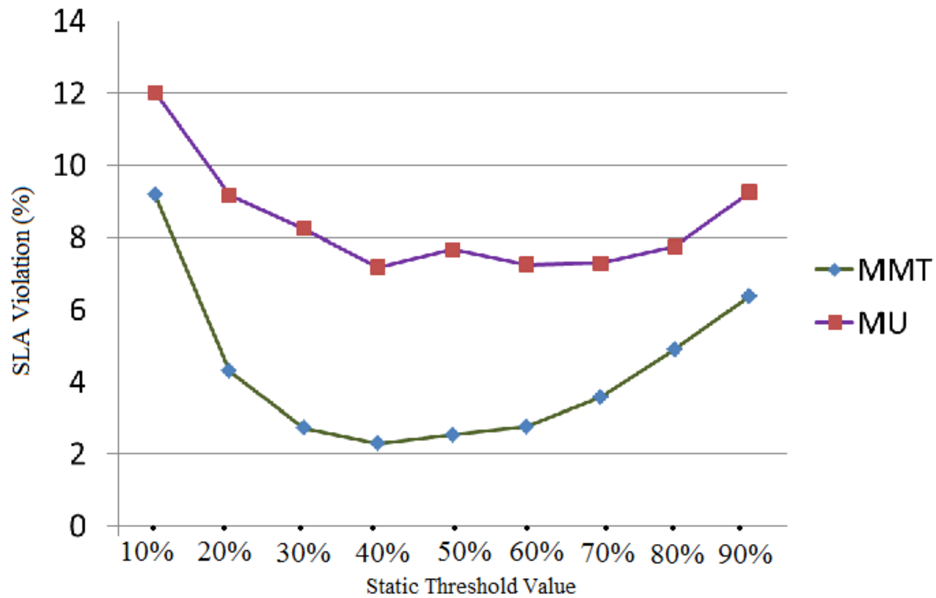

Figure 6: SLA violation results (MMT vs.MU)

From figure 6, the violation (%) to the SLA caused by MU algorithm is very high in comparison to MMT. For both algorithms, the relationship type between the SLA violation and the Threshold value seems to change after exceeding the value “40%” to become a strong proportional relationship. It is also noticed that both algorithms, after the threshold value “40%”, become more prone to violate their SLA and both algorithms appear in their best level regarding SLA at the threshold value “40%”. However, by comparing both graphs for power consumption and SLA violations, it is noticed that the best trade-off between power efficiency and QoS is achieved at threshold value “50%” for both algorithms and the MMT algorithm provided better overall results than MU.

5. Conclusions and Future Work

This work contributes to advancing the cloud computing infrastructure field in two different aspects. Primarily, it significantly contributes toward the reduction of power consumption in cloud-based infrastructure by proposing feasible trade-offs between power consumption and performance. Secondly, two policies have been explored, studied, and evaluated in order to investigate their abilities to contribute towards power consumption reduction with respect to the quality of service issue. This contributes to formulating more sophisticated policies to enrich and strengthen the managerial platforms for such infrastructures.

We have proposed a consolidation approach for reducing the power consumed in cloud datacenters. Two virtual machines selection policies were studied and evaluated to find the more efficient one for such approach. The MMT policy with an optimal threshold value appeared to produce more efficient trade-off between power efficiency and QoS than MU algorithm. The proposed Static Threshold Value (STV) is a key contribution of this paper. Our simulation results have shown that the consolidation policies leads to more power-efficient cloud computing infrastructure in comparison to traditionally operated datacenter when no power-aware optimization techniques are employed. In our future work, we aim at advancing the current approaches to handle complex cases such as more dynamic demands and various QoS requirements by multiple users.

References

Agrawal, D., Lee, K.-W., & Lobo, J. (2005). Policy-Based Management of Networked Computing Systems. *IEEE Communications Magazine*, 69-75.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., et al. (2010). A View of Cloud Computing . *communications of the ac m*, 50-58.

Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., et al. (2003). Proceedings of the 19th ACM Symposium on Operating Systems PrinciplesXen and the Art of Virtualization. Bolton Landing.

Beloglazov, A., & Buyya, R. (2011). Optimal online deterministic algorithms and adaptive heuristics forenergy and performance efficient dynamic consolidation of virtual machines in Cloud data centers. *Concurrency and Computation: Practice and Experience*, 1397–1420.

Beltran, F., Gutierrez, J. and Melus, J. L., "Technology and market conditions towards a new competitive landscape in the wireless access market", IEEE Communications Magazine, Vol. 48, No. 6, 46-52, 2010

Cisco. (2008). Managed Services: Taking Advantage of Managed Services in the High-End Enterprise. *2008 Cisco Systems*, 1-15.

Clark, C., Fraser, K., Hand, S., & Hansen, J. G. (2005). Live Migration of Virtual Machines. *2nd Symposium on Networked Systems Design & Implementation*, 273-286.

Gens, F. (2008, October 08). *IT Cloud Services Forecast*. Retrieved October 23, 2012, from IDC exchange : http://blogs.idc.com/ie/?p=224

Greenpeace. (2010). *Cloud Computing and its Contribution to Climate Change* . Amsterdam: Greenpeace International .

Gutierrez, J.A. “Policy-based Network Management as a platform for the pricing of ubiquitous network services”, Proceedings of the 2008 IEEE International Conference on Intelligent Pervasive Computing (IPC 08), Sydney, Australia, December, 439-443, 2008

Hadjiantonis, A. M., Charalambides, M., & Pavlou, G. (2007). A policy-based approach for managing ubiquitous networks in urban spaces . *IEEE Communications* , (pp. 2089-2096).

Koomey, J. G. (2011). *A report by Analytics Press, completed at the request of The New York Times.* USA : Stanford University.

Kuribayashi, S.-i. (2012). Reducing Total Power Consumption Method in Cloud Computing Environments. *International Journal of Computer Networks & Communications*, 69-84.

McKendrick, J. (2011, March 10). *Cloud Computing's Hidden 'Green' Benefits*. Retrieved October 24, 2012, from Forbes: http://www.forbes.com/sites/joemckendrick/2011/10/03/cloud-computings-hidden-green-benefits/

Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing.* Gaithersburg: NIST, US.

Mishra, M., Das, A., Kularni, P., & Sahoo, A. (2012). Dynamic Resource Management Using Virtual Machine Migrations. *Cloud Computing: Networking and communication challenges.* IEEE Communication Magazine.

Padala, P., Zhu, X., Wang, Z., Singhal, S., & Shin, K. G. (2007). Performance evaluation of virtualization technologies for server consolidation. *Enterprise Systems and Software Laboratory, HP Laboratories Technical*, 1-14.

Ray, S.K. and Gutierrez, J. "On Applications and Profiles for Policy-based Pricing of Ubiquitous Services". Proceedings of the 25th Australasian Conference on Information Systems, Auckland, New Zealand, December, 2014

Rivoire, S., Shah, M. A., Ranganathan, P., & Kozyrakis, C. (2007). Joulesort: a Balanced Energy-Efficiency Benchmark. *ACM SIGMOD International Conference on Management of Data,*.

Rodrigo N. Calheiros, R. R., Beloglazov, A., Rose, C. A., & Buyya, R. (2011). CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms. *Practice and Experience (SPE)*, 23-50.

Rouse, M. (2010, December). *cloud computing*. Retrieved October 26, 2012, from SearchCloudComputing : http://searchcloudcomputing.techtarget.com/definition/cloud-computing

Spellmann, A., Erickson, K., & Reynolds, J. (2003). Server Consolidation Using Performance Modeling. *IT PRO, IEEE*, 31-36.

Verma, D. C. (2002). Simplifying Network Administration Using Policy-Based Management. *IBM Thomas J Watson Research Cente, IEEE Network* , 20-26.

Vouk, M. A. (2008). Cloud Computing -Issues, Research and Implementations . *Journal of Computing and Information Technology* , 235-246.

Zyga, L. (2010). *How energy-efficient is cloud computing?* Retrieved August 11, 2012, from phys.org: http://phys.org/news205737760.html