Survey on Cloud Computing: Energy Saver

Madhuri Gokhale
Assistant Professor, Dept of Computer Science and Engineering
Jabalpur Engineering College, Jabalpur, Madhya Pradesh

Abstract- *Energy efficiency is increasingly important for future information and communication technologies (ICT), because the increased usage of ICT, together with increasing energy costs and the need to reduce greenhouse gas emissions call for energy-efficient technologies that decrease the overall energy consumption of computation, storage, and communications. Cloud computing has recently received considerable attention, as a promising approach for delivering ICT services by improving the utilization of data center resources. In principle, cloud computing can be an inherently energy-efficient technology for ICT provided that its potential for significant energy savings that have so far focused on hardware aspects, can be fully explored concerning system operation and networking aspects. Thus this paper, in the context of cloud computing, reviews the usage of methods and technologies currently used for energy-efficient operation of computer hardware and network infrastructure. After surveying some of the current best practices and relevant literature in this area, this paper identifies some of the remaining key research challenges that arise when such energy-saving techniques are extended for use in cloud computing environments.*

Keywords- Cloud Service Models, Power dissipation, Geo-graphic Adaptive Fidelity Protocol (GAF), Cluster Based Energy Conservation Protocol (CEC)

I. INTRODUCTION

In the last years, cloud computing has become more and more popular. This increase in popularity of cloud services results in higher resource demands on the provider's end. More resources mean more energy consumption and thus higher electricity bills. There is a need to make a cloud service more profitable by reducing energy usage while at the same time keeping the service level for the customer. In this paper, we want to discuss several ways found in scientific literature to achieve this goal. The easiest and most obvious way to save energy is to run fewer machines. This however comes with a trade-off. Customers expect the cloud to handle sudden increases in demand; as such it is not as simple as turning unused machines off. As result algorithms are needed to decide in a smart way which machines can be turned off and if new machines should be powered [1]. Moreover, to get empty machines to turn off, scheduling needs energy awareness.

Since most modern machines have different power states, it is also an option to power on/off not directly but by passing through these different states. By running machines in a low power state, they need less energy but do not provide their full capacity. But the change from one power state to its next is much faster than from full capacity to turn off, so sudden increases can easier be served [2].

II. CLOUD SERVICE MODELS

We focus our attention on three cloud services - storage as a service, processing as a service, and software as a service.

In the following sections, we outline the functionality of each of the three cloud services. Note that we use the terms 'client', 'user', and 'customer' interchangeably.

A. Software as a Service

Consumer software is traditionally purchased with a fixed upfront payment for a license and a copy of the software on appropriate media. This software license typically only permits the user to install the software on one computer. When a major update is applied to the software and a new versions released, users are required to make a further payment to use the new version of the software. Users can continue to use an older version, but once a new version of the software has been released, support for older versions is often significantly reduced and updates are infrequent [3], [4], [5].

With the ubiquitous availability of broadband Internet, soft- ware developers are increasingly moving towards providing software as a service. In this service, clients are charged a monthly or yearly fee for access to the latest version of the software. Additionally, the software is hosted in the cloud and all computation is performed in the cloud. The client's PC is only used to transmit commands and receive results. Typically, users are free to use any computer connected to the Internet. However, at any time, only a fixed number of instances of the software are permitted to be running per user. One example of software as a service is Google Docs.

When a user exclusively uses network- or Internet-based software services, the concept is similar to a “thin client” model, where each user’s client computer functions primarily as a network terminal, performing input, output, and display tasks, while data are stored and processed on a central server.

B. Storage as a Service

Through storage as a service, users can outsource their data storage requirements to the cloud. All processing is performed on the user’s personal computers, which may have only a solid- state drive (e.g., flash-based solid-state storage), and the user’s primary data storage is in the cloud [6], [7], [8]. Data files may include documents, photographs, or videos. Files stored in the cloud can be accessed from any computer with an Internet connection at any time. However, to modify a file, it must first be downloaded, edited using the user’s PC, and then the modified file uploaded back to the cloud. The cloud service provider ensures there is sufficient free space in the cloud and also manages the backup of data. Also, after a user uploads a file to the cloud, the user can grant read and/or modification privileges to other users. One example of storage as a service is the Amazon Simple Storage service.

Fig. 1. The SETI@home Closet

C. Processing as a Service

Processing as a service provides users with the resources of a powerful server for specific large computational tasks. The majority of tasks, which are not computationally demanding, are carried out on the user’s PC. More demanding computing tasks are uploaded to the cloud, processed in the cloud, and the results are returned to the user. Similar to the storage service, the processing service can be accessed from any computer connected to the Internet. One example of processing as a service is the Amazon Elastic Compute Cloud service.

When utilizing a processing service, the user’s PC still performs many small tasks and is consequently required to be more powerful than the “thin client ”considered in the software service [9].

III. IS CLOUD COMPUTING ALWAYS GREENER?

As a growing number of small- and medium-sized organizations (SMOs), such as private companies, hospitals, government agencies, and educational institutions, seek to improve the energy efficiency of their Information Technology (IT) operations by moving computing applications to an Internet-based “cloud” platform, it is becoming increasingly important to understand the associated energy and climate impacts [10]. Until now there was no independent analysis to establish whether this system of Internet-based shared servers for multiple customers is indeed the most eco-friendly choice. To uncover the major factors determining how on- premise server rooms and cloud computing stack up in carbon emissions and energy savings, the Natural Resources Defense Council and WSP Environment Energy have partnered on groundbreaking research, examining five different scenarios with the goal of making it easier for companies to compare options and consider sustainability in their decision-making.

While cloud computing is generally more energy-efficient and has a smaller carbon footprint than on-premise server rooms, not all clouds are created equal: some clouds are greener than others.

An on-premise server room that implements energy efficiency best practices can be a greener alternative than a ‘brown cloud’. Ultimately, the study revealed that various solutions exist for SMOs looking to significantly cut energy waste and emissions from both on-premise server rooms and the ‘cloud’.

IV. ENERGY SAVING MODEL FOR CLOUD COMPUTING

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 [11].

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 [12]. According to ICT energy estimates in the radio access network consumes a major part of the total energy in infrastructure and the cost incurred on energy consumption is sometimes compared with the total cost spent on personnel employed for network operations and maintenance. Jia et al. provided a thorough study on routing protocols for saving energy consumption in sensor networks and wireless networks. Topology Control Protocols such as Geographic Adaptive Fidelity (GAF) and Cluster-Based Energy Conservation (CEC) were also presented.

Micro Sensor Architecture comprises 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. To achieve energy savings, sensors need to put in the sleeping state as the other three states consume a considerable amount of energy. GAF and CEC protocols identify redundant nodes and turn off them to conserve energy.

1) Geographic Adaptive Fidelity Protocol: In the GAF protocol, equivalent nodes are found out by using their geographical information, and then their radios are turned off which saves energy [13]. However, for communication between a pair of nodes, equal nodes 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 the GAF protocol always switch among one of the three states – sleeping, discovery, and active. Initially, a node is in a discovery state with its radio turned on and it exchanges messages with its neighbors. A node inactive state and discovery state can switch to a sleeping state whenever it finds an equivalent node that can perform routing.

2) Cluster Based Energy Conservation Protocol: One of the disadvantages of the GAF protocol is that it needs global location information which may not be available every time [14]. Also, it is very conservative because it guesses its connectivity instead of directly measuring it which leads to fewer energy savings. Another protocol namely Cluster-based Energy Conservation overcomes these disadvantages since it is independent of location information and it directly and adaptively measures network connectivity and finds network redundancy more accurately thus saving more energy. CEC operates in the following steps:

(a) Determining Network Redundancy: It operates by organizing nodes into clusters that 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.

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

(b)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 the node ID, its cluster-ID, and its estimated lifetime. A cluster formation begins with the cluster head selection and then the gateway node selection obtained by exchanging discovery messages.

A node having the longest lifetime selects itself as the cluster head and this information is 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 gate- way 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 [15].

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

V. 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 hardware from different vendors. However, there are immense opportunities in designing more power-efficient scheduling algorithms and cluster configuration in the 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 to implementing the hyper graph partitioning approach in the cloud environment.

B. Proposal based upon energy efficient cluster configuration

Energy-efficient cluster configuration cannot be approached as a single methodology. The cluster comprises a processing unit, memory, storage unit, network, and other peripherals. Each component has its 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. The efficient cooling mechanism, energy-aware data bus, latest processing unit, etc., can also help in the reduction of power utilization in clusters.

VI. CONCLUSION

In this paper, we have investigated the need for power consumption and energy efficiency in the cloud computing model. It has been shown that there are few major components of cloud architecture that are responsible for the high amount of power dissipation in the 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, T. Grance, et al., The nist definition of cloud computing (2011).

[2] A. Berl, E. Gelenbe, M. Di Girolamo, G. Giuliani, H. De Meer, M. Q. Dang, K. Pentikousis, Energy-efficient cloud computing, The computer journal 53 (7) (2010) 1045–1051.

[3] Dubey, D. Wagle, Delivering software as a service, The McKinsey Quarterly 6 (2007) (2007) 2007.

[4] W. Sun, K. Zhang, S.-K. Chen, X. Zhang, H. Liang, Software as a service: An integration perspective, in: International Conference on Service-Oriented Computing, Springer, 2007, pp. 558–569

[5] V. Choudhary, Software as a service: Implications for investment in software development, in: 2007 40th Annual Hawaii International Con- ference on System Sciences (HICSS'07), IEEE, 2007, pp. 209a–209a.

[6] K. Jamsa, Cloud computing: SaaS, PaaS, IaaS, virtualization, business models, mobile, security and more, Jones & Bartlett Publishers, 2012.

[7] D. Rani, R. K. Ranjan, A comparative study of saas, paas and iaas in cloud computing, International Journal of Advanced Research in Computer Science and Software Engineering 4 (6) (2014).

[8] M. J. Kavis, Architecting the cloud: design decisions for cloud comput- ing service models (SaaS, PaaS, and IaaS), John Wiley & Sons, 2014.

[9] J. Baliga, R. W. Ayre, K. Hinton, R. S. Tucker, Green cloud computing: Balancing energy in processing, storage, and transport, Proceedings of the IEEE 99 (1) (2010) 149–167.

[10] M. A. Sharkh, A. Shami, An evergreen cloud: Optimizing energy efficiency in heterogeneous cloud computing architectures, Vehicular Communications 9 (2017) 199–210.

[11] A. Banerjee, P. Agrawal, N. C. S. Iyengar, Energy efficiency model for cloud computing, International Journal of Energy, Information and Communications 4 (6) (2013) 29–42.

[12] T. Mastelic, I. Brandic, Recent trends in energy-efficient cloud comput- ing, IEEE Cloud Computing 2 (1) (2015) 40–47.

[13] S. Roychowdhury, C. Patra, Geographic adaptive fidelity and geographic energy aware routing in ad hoc routing, in: international conference, Vol. 1, Citeseer, 2010, pp. 309–313.

[14] A. Fathi, H. Taheri, Enhance topology control protocol (ecec) to con- serve energy based clustering in wireless ad hoc networks, in: 2010 3rd International Conference on Computer Science and Information Technology, Vol. 9, IEEE, 2010, pp. 356–360.

[15] T. A. AlEnawy, H. Aydin, Energy-aware task allocation for rate mono- tonic scheduling, in: 11th IEEE Real Time and Embedded Technology and Applications Symposium, IEEE, 2005, pp. 213–223.