Energy Consumption in Cloud Computing

Sanil. C. Savale
Department Of CS
Gogate-Jogalekar College
Ratnagiri.

Prashant D. Londhe
Department Of CS
Gogate-Jogalekar College
Ratnagiri.

Ketan G. Jogalekar
Department Of CS
Gogate-Jogalekar College
Ratnagiri.

ABSTRACT

Cloud computing is emerging as a critical information communication technology to heavily impact our daily life in the future. We systematically analyze its energy consumption based on types of services and obtain the conditions to facilitate green cloud computing to save overall energy consumption in the related information communication systems. With a tremendously increasing number of mobile devices, green mobile communications would be the foundation of green cloud computing

Keywords

Energy, Cloud

1. INTRODUCTION

The depletion of fossil energy has become one of the major challenges for mankind to sustain the civilization. In addition, overindulgent energy consumption causes over emission of green-house gas, which, according to expert consensus [6], is a root cause for the current global warming. Seeking to replace fossil energy appears to be an attractive idea, but it may take a long time for the alternatives to attain wide deployment and economic efficiency. It is, therefore, imperative for mankind to seek green technologies, i.e., technologies that can reduce energy consumption.

Among all industries, the information communication technology (ICT) industry is arguably responsible for a large portion of the world-wide growth in energy consumption. This is partly attributed to the rapidly increasing number of Internet and mobile ICT devices available across the globe. As Internet has penetrated into our daily lives, cloud computing has emerged as a new kind of "utility" that gets delivered through wired or wireless networks [3]. Although it is widely claimed that cloud computing is "green" because of its better energy efficiency [10, 2, 5], we are going to examine this thesis more carefully via an analytic approach in this paper. We will investigate whether the energy consumption is reduced by the introduction of cloud computing, as well as to determine under what circumstances cloud computing could be green. Hopefully our conclusion here will provide some insights into and help progress toward greener cloud computing.

Like most buzzwords, there are no universally agreed definitions for cloud computing. There are several well-known pioneering papers on defining what cloud computing is[3, 8], whereas more discussions can be found elsewhere [10, 2, 5].

Based on these works, we summarize that cloud computing is a networked computing structure with the following features:

• Consolidation of computing resources via virtualization;
• Maximization of resource utilization through on-demand, real-time provisioning;
• Delivery of computing, including applications, software, platforms, and infrastructures, as services.

In its own jargon, "the cloud" in cloud computing loosely refers to the remote computing resources, usually under centralized management like servers in data centers that provide services to users over internet. It is the paradigm of cloud computing that users move their computation that is traditionally carried out on personal terminals and local servers to the cloud.

In such a new paradigm, it is of crucial importance that the cloud has good energy efficiency. Srikantaiah et al. point out that a major cause of energy inefficiency in data centers is the idle power wasted when servers run at low utilization [10]. They study how to keep servers at high utilization by workload consolidation. Abdel Salam et al. claim that the request arrival rate at servers varies with time [2].

They develop mechanisms to predict future arrival rates from history and estimate the optimal number of servers for a class of arrival rates. Berl et al. review the literature of achieving higher energy efficiency for data centers, grid computing, and distributed computing in general [5]. They suggest solutions along four dimensions: energy-efficient hardware, energy-aware scheduling, power-minimization in server cluster, and power-minimization in wired and wireless networks.

We stress that our work differs from all the above. We focus on the overall impact on energy consumption brought by cloud computing and find out when it is green. To the best of our knowledge, this is the first analytic work that tries to answer this question using a systematic approach.
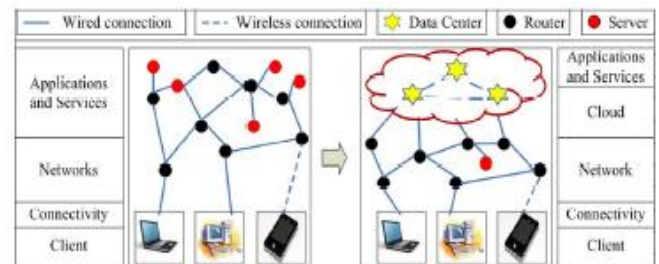

Figure 1: Networking infrastructures with and without the cloud.

2. ANALYTICAL METHODOLOGGY

Figure 1 shows the difference in networking infrastructure brought by cloud computing. The system-wide energy consumption is the sum of energy consumed by all ICT devices involved in the system, which we classify into three categories according to their functionality.

•   Personal terminals: all kinds of personal application devices such as desktop computers, laptop computers, handsets, etc.

•   Networking nodes: communication and networking devices that facilitate connectivity from devices to the cloud, such as routers, switches, hubs, access points, etc.

•   Local servers: the equipment to provide services to personal terminals, such as application servers.

We consider energy consumption induced by computation and communication for personal terminals and local servers, but for networking nodes, we only consider the latter. Based on traffic patterns, we classify applications into three categories as follows.

•   Point and Cloud (Type I): applications that need only communication between user and the cloud. Examples include Dropbox, Remote Desktop, etc.

•   Point-Cloud-Multipoint (Type II): applications in which traffic is initiated from one user to one or more other users through the cloud. Examples include Gmail, Skype, YouTube, etc.

•   Multipoint and Cloud (Type III): applications that provide a platform for users to cooperate on a common piece of work. A good example is Google Docs.

We note that an application's energy consumption should be classified according to its behavior rather than its exact name. For example, if a user uses Google Docs like in an email system and merely sends documents to one or more recipients, then instead of as a Type III application, it should be viewed as a Type II application in this case. The three types of applications are summarized in Figure 2. In that figure, blue and red circles denote personal terminals and local servers, respectively, and yellow stars are data centers in the cloud.
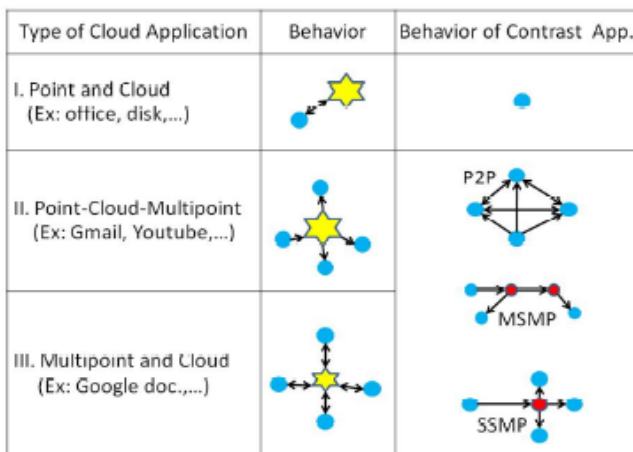

Figure 2: Categorization of cloud applications and their counterparts without the cloud.

The black arrows represent communication links, which also show the direction of information flows. The amount of energy consumed by each type of devices depends heavily on user behavior, which is different with and without the cloud. For example, if a group of users want to collaboratively prepare a set of documents without the cloud, they would have to work on their own personal terminals and exchange the documents through some networking infrastructure. With cloud computing, they can simply collaborate by connecting to the cloud.

In Figure 2, we also list the corresponding contrast applications that can run without the cloud. Obviously, Type I applications are to replace applications on personal terminals, so the contrast applications are those which run completely on personal terminals without communicating with others, e.g., desktop office software. On the other hand, both Type II and Type III applications are to replace client-server or other type of applications that involve some communication. We summarize the three types of contrast applications below.

•   P2P (point-to-point): applications that communicate with each other directly, e.g., BitTorrent.

•   MSMP (multi-server, multi-point): applications that communicate via multiple servers, e.g., email.

•   SSMP (single-server, multi-point): applications that provide a central resource repository, e.g., file transfer.

Finally, we determine and compare the overall system-wide energy consumption with and without cloud computing as follows.

1. For each type of cloud applications and the corresponding contrast applications, we estimate the difference in energy consumption resulted from changes in computation and communication on personal terminals and local servers, as well as in communication on network nodes.

2. We estimate the energy required to operate a cloud, including energy for communication inside and between data centers, energy for computation carried out in data centers, energy for cooling and such, etc. This energy is a fixed energy cost for cloud computing across a wide range of applications.

3. By summing over all devices the energy differences plus the induced operating energy for cloud computing, we come to a conclusion whether the introduction of cloud computing can indeed reduce overall system-wide energy consumption.

3. ENERGY CONSUMPTION ANALYSIS

In this section, we compute the difference in system-wide energy consumption with and without cloud computing to obtain the conditions under which cloud computing is green.

3.1 System Model

Assuming there are Lp categories of personal terminals, Ln categories of network nodes, and Ls categories of servers. The number of devices in the lp-th category of personal terminals is denoted as $N(lp)$; similarly, we have $N(ln)$ and $N(ls)$ for network nodes and servers, respectively. We denote, within a period D, the energy for transmitting one bit and the energy consumed by the communication circuitry of devices in the lp-th category as $Etx(lp)$ and $Ecir(lp)$, respectively; similarly, we have [$Etx(ln);Ecir(ln)$] and [$Etx(ls);Ecir(ls)$] for network nodes and servers, respectively. We denote the numbers of bits transmitted for device m as $Tm(lp)$, $Tm(ln)$, and $Tm(ls)$ for personal terminals, network nodes, and servers, respectively. Then, the overall difference in communication energy consumption can be expressed as:

$$\Delta E(\text{comm}) =$$
$$\Delta \left\{ \sum_{\alpha \in \{p,n,s\}} \sum_{l_\alpha=1}^{L_\alpha} \sum_{m=1}^{N(l_\alpha)} [T_m(l_\alpha)E_{tx}(l_\alpha) + E_{cir}(l_\alpha)] \right\} \quad (1)$$

Within the same period *D*, the computation energy for each category of personal terminals and servers is denoted as *Ecomp(lp)* and *Ecomp(ls)*, respectively. As the computation energy for network nodes remains constant with various traffic loads, the overall difference in computation energy consumption can be expressed as:

$$\Delta E(\text{comp}) = \Delta \left\{ \sum_{\alpha \in \{p,s\}} \sum_{l_\alpha=1}^{L_\alpha} N(l_\alpha)E_{comp}(l_\alpha) \right\}. \quad (2)$$
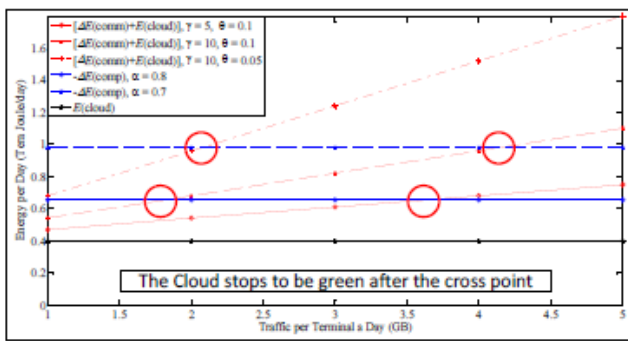

Figure 3: Energy consumption for Type I applications.

Finally, if we denote the energy for operating the cloud as E(cloud), then a necessary condition for green cloud computing is

$$\Delta E(\text{comm}) + \Delta E(\text{comp}) + E(\text{cloud}) < 0: \quad (3)$$

3.2 Energy Analysis For Type I Applications

Figure 2, we can see that the overhead of Type I applications comes mainly from the new traffic they generate. This will certainly increase the communication energy consumption on personal terminals and network nodes, and thus $\Delta E(\text{comm}) > 0$. The computation energy consumption on personal terminals decreases, i.e., $\Delta E(\text{comp}) < 0$, since some computational tasks are offloaded to the cloud. Taking into accounts the cloud operating energy E(cloud), the condition that the overall energy consumption can be reduced can be expressed as:

$$\Delta E(\text{comm}) + E(\text{cloud}) < |\Delta E(\text{comp})| : \quad (4)$$

To better understand the effect of the new traffic, we plot the induced communication energy with traffic due to Type I applications in Figure 3. The interested reader is referred to the appendices for the detailed derivation of $\Delta E(\text{comm})$, E(cloud), and $\Delta E(\text{comp})$. In Figure 3, _ represents then energy saved when personal terminals offload their computation to the cloud. The parameters and _ denote the overhead of the communication network. is the number of bits generated in the network per data bit. Such an overhead may be due to

control signaling like handshaking. _ is the performance degradation ratio for wired and wireless access at personal terminal. For example, even though the theoretical data rate of IEEE 802.11 can be as high as 11 Mbps, in practice it is often a few hundred of Kbps. Such degradation increases the energy cost per bit transmitted, as the power consumption remains high during the entire period of time when the transmitting radio is in operation. From Figure 3, we have the following observations.

• Unlimited traffic would offset, suggesting that data compression and efficient management to reduce traffic is vital to green cloud computing.
• Inefficiency of wired and wireless access significantly increases the energy cost per bit transmitted, suggesting that effective transmission scheme and updated communication infrastructure is vital to green cloud computing.
• It is critical to develop smart traffic routing algorithms in wired and wireless networks in order to reduce traffic overhead that wastes transmission power.
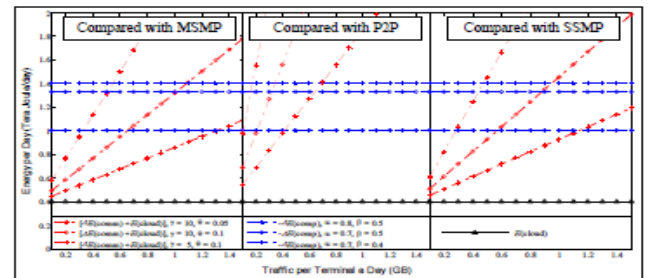

Figure 4: Energy consumption for Type II applications.

3.3 ENERGY ANALYSIS FOR TYPE II APPLICATIONS

The main effect of Type II applications on networking and connectivity is to the changed routing of traffic from local servers to remote servers in the cloud. Therefore, the communication energy of personal terminals is expected to be invariant, while the communication energy of networking nodes are expected to increase. Furthermore, the computation energy consumption on personal terminals and local servers also decreases, as the tasks are offloaded to the cloud. Thus, the increased communication energy consumption due to routing changes needs to be compensated by the net energy saving of moving computation from personal terminals and local servers to the cloud after taking the cloud's fixed operation energy consumption into accounts. The formula for this condition is the same as (4). To delineate the effects of routing changes, Figure 4 compares communication energy consumption for Type II applications. The interested reader is referred to the appendices for more detail of this simulation. Here we only stress the most important parameter _, which denotes the power saving for each server. We observe the following.
• The increased communication energy consumption due to traffic redirection from local to remote servers in the cloud would eventually offset any benefits of computation offloading, no matter compared with MSMP, P2P, or SSMP.
• The computation energy saving from personal terminals is larger than from local server because there are much more personal terminals than local servers. Based on

these observations, we suggest to study energy efficient routing [7], which is not necessarily shortest path routing, as well as distributing energy load evenly throughout the network [9], in order to achieve green cloud computing in this dimension. Moreover, better power management on personal terminals that adapts energy consumption according to computation load is also critical.

3.4 ENERGY ANALYSIS FOR TYPE III APPLICATIONS

Type III applications also change traffic routing from local to remote server in the cloud. The major difference between Type III and Type II applications is that the former provides a platform for users to collaborate on a common task such that instead of the whole data, every user only needs to access the portion that he or she needs to complete his or her part of the work. Therefore, the overall system-wide traffic actually decreases in Type III applications, i.e., both $\Delta E(comm) < 0$ and $\Delta E(comp) < 0$. Thus, the condition under which the cloud computing is green becomes:

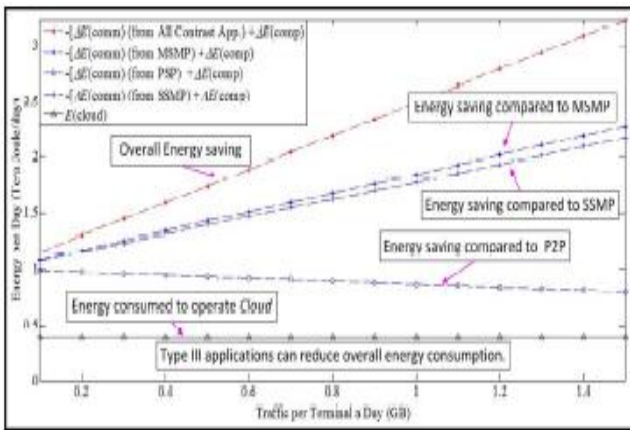Figure 5: Energy consumption for Type III applications.

$$E(cloud) < |\Delta E(comp)| + |j\Delta E(comm)| : \qquad (5)$$

Figure 5 shows the potential energy savings for Type III applications, where we assume _ = 0:8, _ = 0:5, _ = 0:1, and = 5. We can observe that the saved computation and communication energy is in general larger than the energy cost to operate the cloud. Although the energy saving from P2P decreases with traffic, the total energy saving from all contrast applications increases with traffic. Therefore, Type III applications can indeed effectively reduce overall system-wide energy consumption.

3.5 EFFECT OF INCREASING TERMINALS

Major factor for increased ICT energy consumption is the exponentially growing number of terminals, a phenomenon known as the network effect. Such an increasing number of terminals would not only increase the communication energy but also decrease the computation energy saving from existing terminals and servers.

In Figure 6, we plot the trend of communication energy consumption increasing and computation energy consumption decreasing as the number of wireless terminals increases, taking into accounts the cloud operation energy consumption. As long as the per-terminal computation energy saving is smaller than communication energy consumption, the cloud computing cannot support indefinitely many terminals in an energy efficient manner.

Finally, we note that although the increased number of terminals brings new businesses and revenues, it cannot be considered positive from an energy-consumption viewpoint. To conquer this problem, in addition to effective data compression and management, low-power circuit and algorithm design for both computation and communication are critical [7, 4].

4. CONCLUDING REMARKS
In this paper, we have systematically analyzed the energy consumption of cloud computing based on the types of services and obtained the conditions under which the overall system-wide energy consumption is reduced. We point out that data management, compression and efficient network access and infrastructure are critical to facilitate green cloud computing. Furthermore, we suggest studying intended, non-broadcasting routing to control routing overhead, as well as low-power terminal design to mitigate the ever-increasing energy consumption by exponentially growing number of mobile terminals. Based on our observations, we conclude that green mobile communications would be a foundation for green cloud computing.

5. REFERENCES
[1] Energy logic:reducing data center energy consumption by creating savings that cascade across systems, 2009.

[2] H. AbdelSalam, K. Maly, R. Mukkamala, M. Zubair, and D. Kaminsky. Towards energy efficient change management in a cloud computing environment. Scalability of Networks and Services, 5637:161–166, 2009.

[3] M. Armbrust et al. Above the clouds: A berkeley view of cloud computing. Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, February 2009.

[4] L. Benini, A. Bogliolo, and G. Micheli. A survey of design techniques for system-level dynamic power management. IEEE Transactions on VLSI System, 10(3):299–316, June 2000.

[5] A. Berl, E. Gelenbe, M. Girolamo, G. Giuliani, H. Meer, M. Dang, and K. Pentikousis. Energy-efficient cloud computing. The Computer Journal Advance Access, August 2009.

[6] L. Bernstein. Climate change 2007: Synthesis report. Report of the Intergovernmental Panel on Climate Change, 2007.

[7] G. Fettweis, K. Chen, and R. Tafazoli. Green radio: Energy efficiency in wireless networks. KICS/IEEE Journal of Comm. and Networks, 12(2), April 2010.

[8] J. Geelan. Twenty-one experts define cloud computing. Virtualization Journal, January 2009.

[9] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton.Mad skills: New analysis practices for big data. PVLDB,2(2):1481–1492, 2009.

[10] B. F. Cooper, R. Ramakrishnan, U. Srivastava, A. Silberstein, P. Bohannon, H.-A. Jacobsen, N. Puz, D. Weaver, and R. Yerneni. PNUTS: Yahoo!'s hosted data serving platform. Proc. VLDB Endow., 1(2):1277– 1288, 2008.

[11] C. Curino, E. Jones, Y. Zhang, E. Wu, and S. Madden. Relational Cloud: The Case for a Database Service. Technical Report 2010-14, CSAIL, MIT, 2010. http://hdl.handle.net/1721.1/52606.

[12] S. Das, S. Agarwal, D. Agrawal, and A. El Abbadi. ElasTraS: An Elastic, Scalable, and Self Managing Transactional Database for the Cloud. Technical Report 2010-04, CS, UCSB, 2010.

[13] S. Das, D. Agrawal, and A. El Abbadi. ElasTraS: An Elastic Transactional Data Store in the Cloud. In USENIX HotCloud, 2009.

[14] S. Das, D. Agrawal, and A. El Abbadi. G-Store: A Scalable Data Store for Transactional Multi key Access in the Cloud. In ACM SOCC, 2010.

[15] S. Das, S. Nishimura, D. Agrawal, and A. El Abbadi. Live Database Migration for Elasticity in a Multitenant Database for Cloud Platforms. Technical Report 2010-09, CS, UCSB, 2010.

[16] S. Das, Y. Sismanis, K. Beyer, R. Gemulla, P. Haas, And J. McPherson. Ricardo: Integrating R and Hadoop. In SIGMOD,2010.

[17] J. Dean and S. Ghemawat. MapReduce: simplified data Processing on large clusters. In OSDI, pages 137–150, 2004.

[18] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels. Dynamo: Amazon's highly available key-value store. In SOSP, pages 205–220, 2007.

[19] D. J. Dewitt, S. Ghandeharizadeh, D. A. Schneider, A. Bricker, H. I. Hsiao, and R. Rasmussen. The Gamma Database Machine Project. IEEE Trans. on Knowl. And Data Eng., 2(1):44–62, 1990.

[20] The Apache Hadoop Project. http://hadoop.apache.org/core/, 2009.

[21] P. Helland. Life beyond Distributed Transactions: An Apostate's Opinion. In CIDR, pages 132–141, 2007.

[22] D. Jacobs and S. Aulbach. Ruminations on multi-tenant databases. In BTW, pages 514–521, 2007.

[23] T. Kraska, M. Hentschel, G. Alonso, and D. Kossmann. Consistency Rationing in the Cloud: Pay only when it matters. PVLDB, 2(1):253–264, 2009.

[24] D. B. Lomet, A. Fekete, G. Weikum, and M. J. Zwilling. Unbundling transaction services in the cloud. In CIDR Perspectives, 2009.

[25] A. Pavlo, E. Paulson, A. Rasin, D. J. Abadi, D. J. DeWitt, S. Madden, and M. Stonebraker. A comparison of approaches to large-scale data analysis. In SIGMOD, pages 165–178, 2009.

[26] B. Reinwald. Database support for multi-tenant applications. In IEEE Workshop on Information and Software as Services, 2010.

[27] J. B. Rothnie Jr., P. A. Bernstein, S. Fox, N. Goodman, M. Hammer, T. A. Landers, C. L. Reeve, D. W.Shipman, and E. Wong. Introduction to a System for Distributed Databases (SDD-1). ACM Trans. Database Syst., 5(1):1– 17, 1980.

[28] A. Thusoo, J. S. Sarma, N. Jain, Z. Shao, P. Chakka, S. Anthony, H. Liu, P. Wyckoff, and R. Murthy. Hive – A Warehousing Solution Over a Map-Reduce Framework. PVLDB, 2(2):1626–1629, 2009.

[29] H. T. Vo, C. Chen, and B. C. Ooi. Towards elastic transactional cloud storage with range query support. PVLDB, 3(1):506–517, 2010.

[30] W. Vogels. Data access patterns in the amazon.com Technology platform. In VLDB, pages 1–1. VLDB Endowment, 2007.