ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

Using Shortest Job First Scheduling in Green Cloud Computing

Abeer H. El Bakely[1], Hesham A.Hefny[2]

Department of CS and IS, ISSR Cairo University, Cairo, Egypt[1, 2]

Abstract: Researchers try to solve the problem of energy (the demand go up and the supply is declining or flat) by discovering new resources or minimizing energy consumption in most important fields. In this paper we minimize energy in cloud computing by enhancement green scheduler, a cloud datacenter comprises of many hundred or thousands of networked servers. The network system is main component in cloud computing which consumes a non-negligible fraction of the total power consumption. In this paper we minimize energy in cloud computing by presenting a scheduling approach that is called SJFGC which performs the best-effort workload consolidation on a minimum set of servers. The proposed approach optimizes the tradeoff between job consolidations (to minimize the amount of computing servers) by executing firstly task with minimum arrival time and minimum processing time. We use simulator program is called Green Cloud which is extension to network simulator NS2.

Keywords: Green Scheduling, Data center, Green Cloud, Energy Efficiency.

I. INTRODUCTION

Energy is involved in all life cycles, and it is essential in all productive activities such as space heating, water lifting, and hospitals ….. etc, energy demands in the world go up, and energy supply is declining or flat. So there is a big challenge to all researches, they try to decline energy consumption or find new sources for energy especially in the things affect on our life.

In recent 10 years, Internet has been developing very quickly; the services grow largely through an internet such as electronic mail, web hosting, web services and etc. So the world needs new concept of current technology or reuse the current technology so the cloud computing appears as a recent trend in IT that [13] moves computing and data away from desktop and portable PCs or servers into computer clusters (big data centers owned by the cloud service providers).

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services which have long been referred to Software as a Service (SaaS). The datacenter hardware and software is what we will call a cloud. People may be users or providers of SaaS, or users or providers of Utility Computing.

The cloud datacenters are quite different from traditional hosting facilities. A cloud datacenter comprises of many hundreds or thousands of networked servers with their corresponding storage and networking subsystems, power distribution and conditioning equipment, and cooling infrastructure. Due to large number of equipment, data centers can consume massive energy consumption and emit large amount of carbon. [6] The network system is another main component in cloud computing which consumes a non-negligible fraction of the total power consumption. In cloud computing, since resources are accessed through Internet, both applications and data are needed to be transferred to the compute node.

It requires much more data communication bandwidth between user's PC to the cloud resources than require the application execution requirements. In the network infrastructure, the energy consumption depends especially on the power efficiency and awareness of wired network, namely the network equipment or system design, topology design, and network protocol design. Most of the energy in network devices is wasted because they are designed to handle worst case scenario. The energy consumption of these devices remains almost the same during both peak time and idle state. Many improvements are required to get high energy efficiency in these devices. For example during low utilization periods, Ethernet links can be turned off and packets can be routed around them. Further energy savings are possible at the hardware level of the routers through appropriate selection and optimization of the layout of various internal router components (i.e. buffers, links, etc.). [6]

Roughly 40% of energy consumption for the world is related to the energy consumed by information technology (IT) equipment which includes energy consumed by the computing servers as well as data center network hardware used for interconnection. In fact, about one-third of the total IT energy is consumed by communication links, switching, and aggregation elements, while the remaining two-thirds are allocated to computing servers. Other systems contributing to the data center energy consumption are cooling and power distribution systems that account for 45% and 15% of total energy consumption. [9]

There are many solutions that are implemented for making data center hardware energy efficient. There are two common techniques for reducing power consumption in computing systems. The Dynamic Voltage and Frequency Scaling (DVFS) enables processors to run at different combinations of frequencies with voltages to reduce the power consumption of the processor. [15]

Management (DPM) achieves most of energy savings by coordinating and distributing the work between all available nodes.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

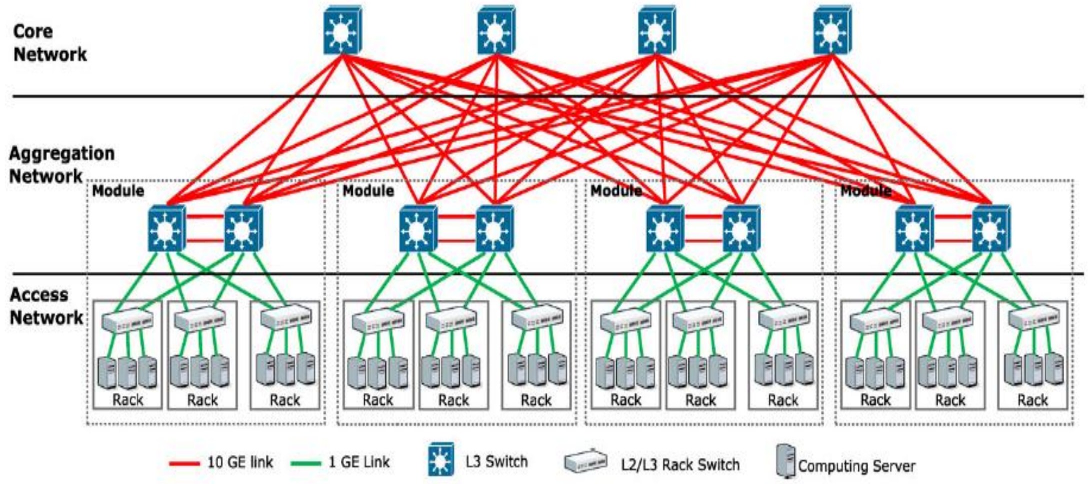

Fig.1GreenCloud simulator architecture (three tiers) [8]

To make DPM scheme efficient, a scheduler must consolidate data center jobs on a minimum set of computing resources to maximize the amount of unloaded servers that can be powered down (or put to sleep).
The average data center workload often stays around 30%, so the portion of unloaded servers can be as high as 70%. [14][9]
Green Cloud simulator is an extension of NS2 which represents the cloud data center's energy efficiency by using two techniques which are DVFS and DPM.
Most of the existing approaches for energy-efficient focus on other targets such as balance between energy efficient and performance by job scheduling in data centers, reduce traffic and congestion in networks of cloud computing. This paper presents a data center scheduling approach which increases an improvement in energy consumption and achieves other advantages such as balance between energy efficient and performance, reduce traffic and congestion in networks of cloud computing. It gets the task with minimum arrival time which toward from user to computing server for execution firstly. The scheduling approach in this paper is designed to minimize the number of computing servers required for job execution.
The proposed approach reduces computational and memory overhead compared to previous approaches, such as flow differentiation, which makes the proposed approach easy to implement and port to existing data center schedulers. Also it reduces complexity time of processing compared to previous approaches.
The rest of the paper is organized as follows: Section 2 presents the related works, Section 3 focuses on environment of simulation, Section 4 describes SJFGC as a proposed approach, Section 5 introduces simulation scenario of proposed approach; Section 6 presents results of proposed approach and Section 7 presents conclusion.

II. RELATEDWORKS

Through reviewing the literature to stand on what other researchers have reached in this research area, a number of subjects of interest were found and can be summarized as follows;

[Dzmitry et al, 2013] proposed e-STAB algorithm for scheduling, it takes into account traffic requirements of cloud applications providing energy efficient job allocation and traffic load balancing in the data center networks. Effective distribution of network traffic improves quality of service of running cloud applications by reducing the communication-related delays and congestion-related packet losses. [11]
[DzmitryKliazovich, 2011] proposes approach (DENS) that balances the energy consumption of a data center, individual job performance, and traffic demands. Methodology aims to achieve the balance between individual job's performance, Quality-of-Service requirements, traffic demands, and energy consumed by the data center. The proposed approach optimizes the tradeoff between job consolidation (to minimize the amount of computing servers) and distribution of traffic patterns (to avoid hotspots in the data center network).
DENS methodology is particularly relevant in data centers running data-intensive jobs which require low computational load, but produce heavy data streams directed to the end-users. [9]
[Haiyang Qian, 2012] has proposed aggregation workload demand to minimize server operational cost by leveraging switching servers on/off and DVFS when resource requirements (workload demand) are given. The numerical results show that aggregation workload over time slots is efficient with small overhead for dynamic aggregation method. It is worked on decomposing workload demand into special substructures and developing algorithms to compute optimal solutions according to the substructures. The substructure based algorithms are more efficient than the mathematical programming based method. [5]

III. ENVIRONMENT OF SIMULATION

We use Green Cloud simulator which is an extension to the network simulator NS2 which is developed for the study of cloud computing environments. The Green Cloud offers users a detailed fine-grained modelling of the energy consumed by the elements of the data center, such as servers, switches, and links. Moreover, Green Cloud

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

offers a thorough investigation of workload distributions. Furthermore, a specific focus is devoted on the packet-level simulations of communications in the data center infrastructure, which provides the finest-grain control and is not present in any cloud computing simulation environment. [8]

The Green Cloud simulator implements energy model of switches and links according to the values of power consumption for different elements. The implemented powers saving schemes are: (a) DVFS only, (b) DNS only, and (c) DVFS with DNS. [8]

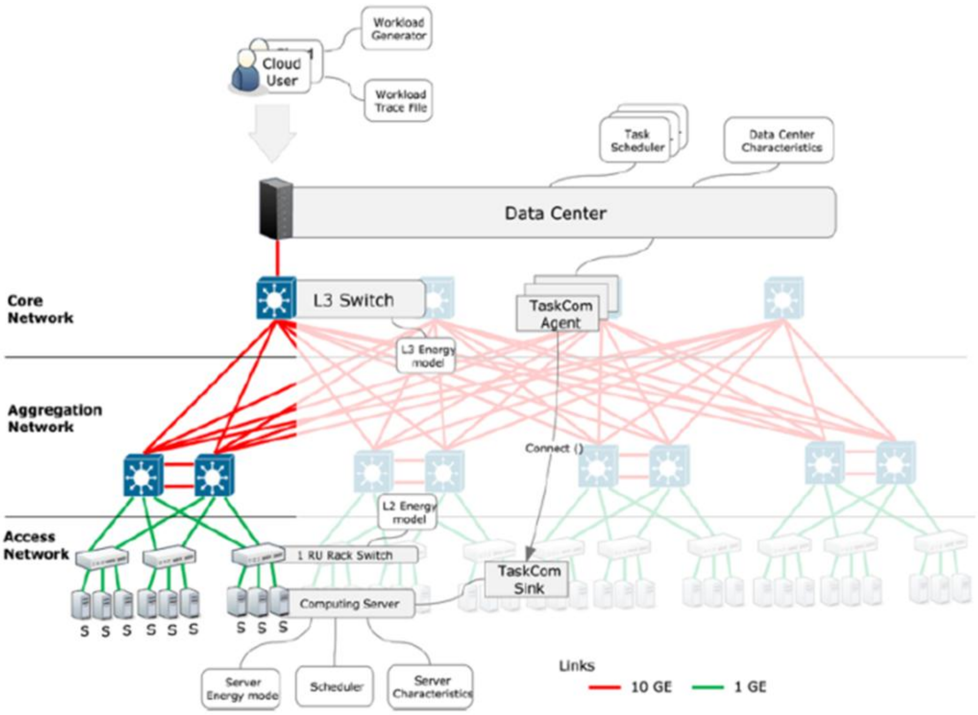

Fig. 2 Architecture of the Green Cloud simulation environment [9]

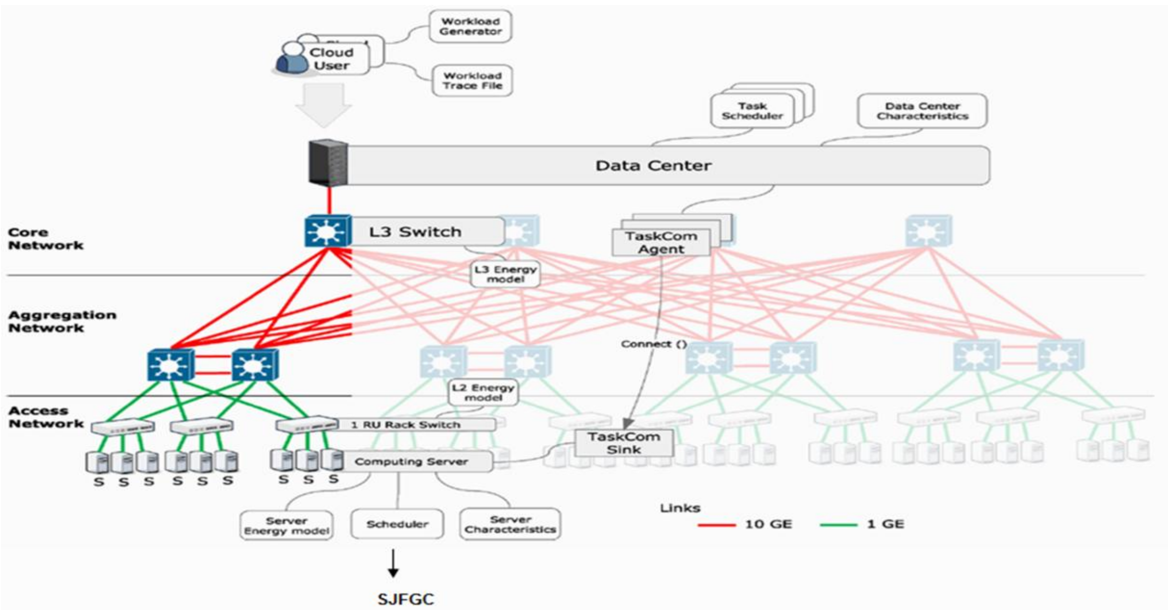

Fig.3 Green cloud Architecture with SJFGC scheduling

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

A. Data Center Topology

Three-tier trees of hosts and switches form which is most common used as data center architecture.
It (see Fig. 2) includes: access, aggregation and core layers. The core tier at the root of the tree, the aggregation tier is responsible for routing, and the access tier that holds the pool of computing servers (or hosts).
The availability of the aggregation layer facilitates the increase in the number of server nodes while keeping inexpensive Layer-2 (L2) switches in the access network, which provides a loop-free topology. Because the maximum number of Equal Cost Multi-Path (ECMP) paths allowed is eight, a typical three tier architecture consists of eight core switches. Such architecture implements an 8-way ECMP that includes 10 GE Line Aggregation Groups (LAGs), which allow a network client to address several links and network ports with a single MAC (Media Access Control) Address. [9][16]
In three-tier architecture the computing servers (grouped in racks) are interconnected using 1 Gigabit Ethernet (GE) links. The maximum number of allowable ECMP paths bounds the total number of core switches to eight. Such a bound also limits the deliverable bandwidth to the aggregation switches. At the higher layers of hierarchy, the racks are arranged in modules (see Fig. 1) with a pair of aggregation switches servicing the module connectivity.
The bandwidth between the core and aggregation networks is distributed using a multi-path routing technology, such as ECMP routing. The ECMP technique performs a per-flow load balancing, which differentiates the flows by computing a hash function on the incoming packet headers. [9]

B. Simulator Components

Computing servers are basic of data center that are responsible for task execution, so it is main factor in energy consumption.
In Green Cloud, the server components implement single core nodes that have a preset on a processing power limit in MIPS or FLOPS, associated size of the memory resources, the power consumption of a computing server is proportional to the CPU utilization. An idle server consumes around two-thirds of its peak-load consumption to keep memory, disks, and I/O resources running. The remaining one-third changes almost linearly with the increase in the level of CPU load.
There are two main approaches for reducing energy consumption in computing servers: (a) DVFS and (b) DPM. The DVFS scheme adjusts the CPU power according to the offered load. The fact that power in a chip decreases proportionally to V 2 *f, where V is a voltage, and f is the operating frequency. This implies a cubic relationship from f in the CPU power consumption. The scope of the DVFS optimization is limited to CPUs. Computing server components, such as buses, memory, and disks remain functioning at the original operating frequency.
The DPM scheme can reduce power of computing servers (that consist of all components); the power model followed by server components is dependent on the server state and its CPU utilization. An idle server consumes about 66% of its fully loaded configuration. This is due to the fact that servers must manage memory modules, disks, I/O resources, and other peripherals in an acceptable state. Then, the power consumption increases with the level of CPU load linearly. Power model allows implementation of power saving in a centralized scheduler that can provision the consolidation of workloads in a minimum possible amount of the computing servers. [9][10][16]

Switches and Links form the interconnection fabric that delivers job requests and workload to any of the computing servers for execution in a timely manner.
The interconnection of switches and servers requires different cabling solutions depending on the supported bandwidth, physical and quality characteristics of the link. The quality of signal transmission in a given cable determines a tradeoff between the transmission rate and the link distance, which are the factors defining the cost and energy consumption of the transceivers.
Energy consumption of a switch depends on the:
(a) Type of switch, (b) Number of ports, (c) Port transmission rates and (d) Employed cabling solutions.
The energy is consumed by a switch can be generalized by the following:

$$P_{\text{switch}} = P_{\text{chassis}} + n_{\text{linecards}} + P_{\text{linecard}} + \sum_{i=0}^{R} n_{\text{ports},r} + P_r \quad \textbf{(1)[9]}$$

Where $P_{chassis}$ is related to the power consumed by the switch hardware, $P_{linecard}$ is the power consumed by any active network line card, P_r corresponds to the power consumed by all of the switches can dynamically be put to sleep. Each core switch consumes a certain amount of energy to service large switching capacity. Because of their location within the communication fabric and proper ECMP forwarding functionality, it is advisable to keep the core network switches running continuously at their maximum transmission rates. On the contrary, the aggregation switches service modules, which can be reduced energy consumption when the module racks are inactive. The fact that on average most of the data centers are utilized around 30% of their compute capacity, it shows power down of unused aggregation switches. However, such an operation must be performed carefully by considering possible fluctuations in job arrival rates. Typically, it is enough to keep a few computing servers running idle on top of the necessary computing servers as a buffer to account for possible data center load fluctuation. [10]

IV. SJFGC APPROACH

In SJFGC approach, we introduce the idea of shortest-Job-First scheduling in operating system, shortest-Job-First scheduling associates with each process the length of the latter's next CPU burst. When the CPU is available, it is assigned to the process that has the smallest next CPU burst. The next CPU burst is generally predicted as an exponential average of the measured lengths of previous CPU bursts. [12]

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

In this approach, it is enhanced the green scheduler which performs the best-effort workload consolidation on a minimum set of servers in green cloud computing by implementing function to execute firstly task with minimum arrival time (exponential distribution).

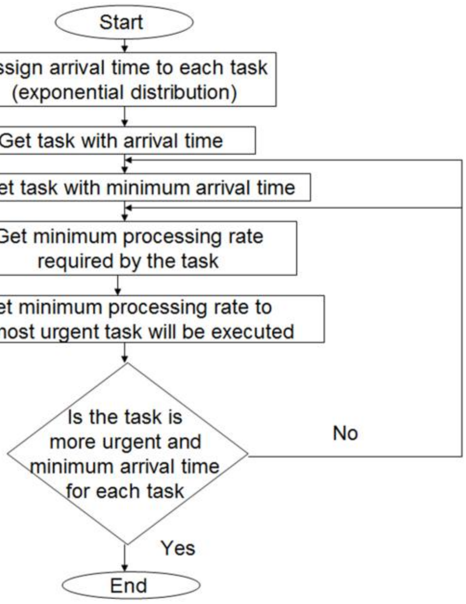

Fig.4 enhancing green scheduling algorithm in green cloud by implementing function

This chart describes implementation of enhancement to green scheduler by add set of functions and processes for selection the task has minimum arrival time:

- Assign arrival time to each task which follows exponential distribution: Assuming arrival time follows exponential distribution from NS2 then assigning this variable to task class.
- Get task with arrival time: this function is implemented to get the task with arrival time.
- Get task with minimum arrival time: this function is implemented in host class to get the task with minimum arrival time.
- Get minimum processing rate required by the task: this function is implemented in host class in green scheduler to get minimum processing rate.
- Get minimum processing rate by the most urgent task will be executed: this function is implemented in green scheduler to get the minimum processing rate by the most urgent already executed task and we add in this function calling to function “get minimum arrival time”
- Check whether the task is more urgent and has minimum arrival time: this question to be sure the task will be executed will be more urgent and has minimum arrival time. If the answer is yes the task will executed and if the answer is no it will revaluate the arrival time and the urgent of task.

V. SIMULATION SCENARIO

A three-tier tree data center topology comprised of 1536 servers arranged into 32 racks each holding 48 servers, served by 4 core and 8 aggregation switches (see Fig. 2),it is used in simulation experiment. We used 1 GE links for interconnecting servers in the inside racks while 10 GE links were used to form a fat-tree topology interconnecting access, aggregation and core switches. The size of the workload is equal to 15 KB. Being fragmented, it occupies 10 Ethernet packets.During executions; the workloads produce a constant bitrate stream of 1 Mb/s directed out of the data center.

Such a stream is designed to mimic the behavior of the most common video sharing applications. To add uncertainties, during the execution, each workload communicates with another randomly chosen workload by sending a 75 KB message internally. The message of the same size is also sent out of the data center at the moment of task completion as an external communication. [8, 9]

TABLE I. SIMULATION SETUP PARAMETERS

	Parameter	**Data Center Architectures**
Topologies	Core nodes (C1)	8
	Aggregation nodes (C2)	16
	Access switches (C3)	512
	Servers (S)	1536
	Link (C1–C2)	10 GE
	Link (C2–C3)	1GE
	Link (C3–S)	1 GE
Data Center	Data center average load	30%
	Task generation time	Exponentially distributed
	Task size	Exponentially distributed
	Simulation time	60 minutes

The workload generation events are exponentially distributed in time to mimic typical process of user arrival. As soon as a scheduling decision is taken for a newly arrived workload it is sent over the data center network to the selected server for execution. The propagation delay on all of the links was set to 10 ns.

The server peak consumption is 301 W which is composed between 130W allocated for a peak CPU consumption and 171 W is consumed by other devices. The minimum consumption of an idle server is 198W.

The average load of the data center is kept at 30% that is distributed among the servers using two scheduling schemes: (a) SJFGC scheduling approach is proposed in Sec. 4 of this paper, theswitches consumption is almost constant for different transmissionrates because the most of thepower is consumed by their chassis and line cards and only a small portion is consumed by their port transceivers. For the 3T topology where links are 10 G the core while 1 G aggregation and rack.(b) Green schedulertends to group the workload on a minimum possible amount of computing servers. [10]

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

VI. RESULTS OF SIMULATION

In compared approach, the workloads arrived to the data center and are scheduled for execution using energy aware "green" scheduler. This "green" scheduler tends to group the workloads on a minimum possible amount of computing servers. The scheduler continuously tracks buffer occupancy of network switches on the path. In case of congestion, the scheduler avoids using congested routes even if they lead to the servers able to satisfy computational requirement of the workloads.
The servers left idle are put into sleep mode (DNS scheme), the time required to change the power state in either mode is set to 100 ms. [9]

TABLE II. COMPARISON OF ENERGY-EFFICIENT OF GREEN SCHEDULER AND SJFGC APPROACH

	Energy Consumption (KW h)		
	Compared [8]	**SJFGC**	**An Improvement of energy**
Server	138.4	91.35442	33.99%
Switch	48.3	24.32081	49.65%
Data center	186.7	115.6752	38.04%

In simulation work, we use DNS scheme for minimizing energy consumption, in compared work we use green scheduler, and in proposed work we use SJFGC approach, we enhance the green schedulerby execution firstly task with minimum arrival time which is followed exponential distribution.
We measure an improvement of energy as follows:
An Improvement of energy = (1- (SJFGC /green scheduler))*100 (2)
Selection the server to compute selected task is randomly so the results of table (1) is average of 20 runs.
In this paper, Table 2 represents comparison between different scheduling; the data is collected for an average data center load of 30%. In applied SJFGC approach on DNS scheme the energy consumption is reduced in server, switch and data center.
Figure 6 shows an improvement in server energy is 33.99%, in switch energy is 49.65% and in data center is 38.04%. An improvement in energy consumption of switch is better than server and data center, this means that using SJFGC minimizes energy consumption in all components of cloud computing because it is more effect on computing server.
All improvements in server, switch and data center mean SJFGC approach is better than green scheduler approach.

VII. CONCLUSION

In this paper, we present a proposed approach to minimize energy in cloud computing, which is based on enhancing green scheduler by using the idea of shortest job first scheduling in operating system..
In SJFGC approach, execution firstly task with shortest arrival time to computing server which is followed exponential distribution.

From Comparison between SJFGC and green scheduler, we conclude the proposed approach is better than the compared approach especially in energy consumption in switch because the percentage of improvement is better than server and data center.

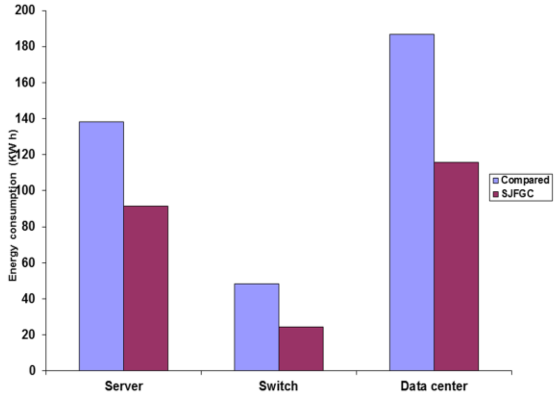

Fig. 5 Comparison between green scheduler and SJFGC of energy consumption

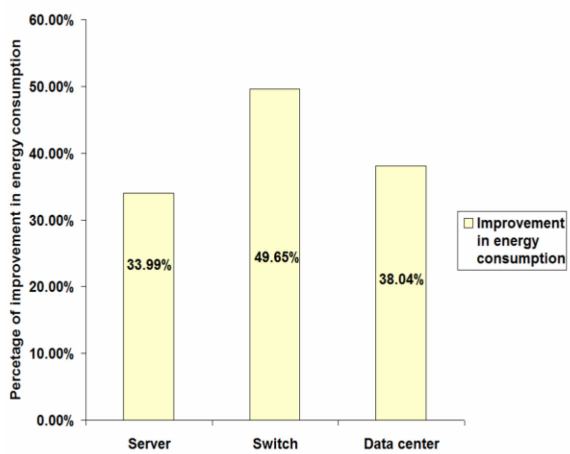

Fig.6 percentage of an improvement in energy consumption

At result, the proposed approach is better than compared approach in minimizing energy consumption of cloud computing.

REFERENCES

[1] V.Kumar, "Temperature-Aware Virtual Machine Scheduling in Green Clouds", M. thesis, Univ. of Thapar, Patiala, Jun. 2012
[2] L.Ganesh, "DATACENTER ENERGY MANAGEMENT", "D. thesis", Univ. of Cornell, Jan. 2012
[3] D.Mann, "A Heterogeneous Workload Consolidation Technique for Green Cloud", "M. thesis", Univ. of Thapar, Jun. 2012
[4] A.J. Younge, "TOWARDS A GREEN FRAMEWORK FOR CLOUD DATA CENTERS", "M. thesis", Rochester Institute of Technology, May. 2010
[5] H.Qian, "DATA CENTER RESOURCE MANAGEMENT WITH TEMPORAL DYNAMIC WORKLOAD", "D. thesis", Univ. of Missouri–Kansas City, 2012
[6] S.Kumar Garg , R.Buyya, "Green Cloud computing and Environmental Sustainability", Dept. of Computer Science and Software Engineering,Univ. of Melbourne, Australia,2011
[7] F.Seng Chu, K.Cheng Chen, C.Mou Cheng, "Toward Green Cloud Computing", ICUIMC, February 21-23, 2011, Seoul, korea,ACM,2011

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

[8] D.Kliazovich, P.Bouvry, S.U.Khan, "GreenCloud: a packet-level simulator of energy-aware cloud computing data centers", Springer Science+Business Media, 9 Nov. 2010

[9] D.Kliazovich, P.Bouvry, S.U.Khan," DENS: data center energy-efficient network-aware scheduling", Springer Science+Business Media, 10 Aug. 2011

[10] D.Kliazovich, P.Bouvry, S.U.Khan, "GreenCloud: A Packet-level Simulator ofEnergy-aware Cloud Computing Data Centers",in proc.IEEE Globecom, 2010,paper 978-1-4244-5637-6/10

[11] D.Kliazovich1, S.T. Arzo, F.Granelli, P.Bouvry ,S.U.Khan3, "e-STAB: Energy-Efficient Scheduling for Cloud Computing Applications with Traffic Load Balancing", IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, 2013

[12] B.S.Gill, S.k.Gill, P.Jain, "Analysis of Energy Aware Data Center using Green Cloud Simulator in Cloud Computing", International Journal of Computer Trends and Technology (IJCTT), vol. 5 number 3, Nov. 2013

[13] A.Silberschatz, P.B.Galvin, Operating system concepts, 5th ed,USA, Addison Wesley Longman, 1998

[14] R.Satsangi, P.Dashore and N.Dubey, "Risk Management in Cloud Computing Through Fuzzy Logic" , IJAIEM, Vol.1, Issue 4, Dec. 2012

[15] Wissam.C, Chansu.Y, "Survey on Power Management Techniques for Energy Efficient Computer Systems", ClevelandStateUniversity, 2003.

[16] Chia.-M.W, Ruay.-S.C, Hsin.-Y.C, "A green energy-efficient scheduling algorithm using DVFS technique for cloud datacenters", Future Generation Computer Systems, 2013, http://dx.doi.org/10.1016/j.future.2013.06.009