A MORE REALISTIC SIMULATION OF PLANETLAB WORKLOAD USING DISSECT-CF

Hasanein D. Rjeib[1], Gabor Kecskemeti[2]
[1]PhD student, PhD, [2]Associate Professor
[1,2]Institute of Information Technology, Faculty of Mechanical Engineering and Informatics, [1]University of Kufa, ECE Department, College of Engineering
[1]rjeib.hasanein.dakheelallah.habeeb@student.uni-miskolc.hu,
[2]kecskemeti@iit.uni-miskolc.hu

ABSTRACT

Cloud computing has become an indispensable part of modern IT infrastructure, offering scalable and on-demand resources to users. As the demand for cloud services continues to grow, researchers and practitioners are faced with the challenge of understanding and optimizing the energy consumption of data centers, as well as accurately modeling the various components within these dynamic environments. In this paper, we explore the importance of realistic simulation in cloud computing, with a focus on the accuracy of energy consumption calculations and the support for various states of data center components.

1. INTRODUCTION

Cloud data centers play a pivotal role in modern computing, providing scalable and efficient infrastructure for hosting applications and services. Realistic simulation of these data centers is crucial for understanding their performance, optimizing resource utilization, and addressing environmental concerns such as energy consumption. Simulating energy consumption accurately requires a comprehensive energy model that encompasses various aspects, including servers and networking equipment. For instance, modeling the dynamic power consumption of servers based on their utilization levels and accounting for the static power draw is crucial for capturing the nuances of energy usage in a data center.

Moreover, considering the impact of workload variations on energy consumption is necessary for a realistic simulation. A robust energy model should integrate workload-aware power models, reflecting the fluctuating demands on the data center resources. This enables the simulation to mirror real-world scenarios where energy consumption varies based on the nature and intensity of workloads. To enhance accuracy, researchers often leverage trace-based simulations, where the simulation parameters and workload characteristics are derived from real-world traces. Additionally, continuous refinement of simulation models based on empirical data helps in minimizing the gap between simulated and actual performance.

The integration of the algorithm proposed by Beloglazov et al. [1] and the utilization of the PlanetLab [2] experiment within CloudSim [3] have collectively established a standard benchmark in the field. This integration has catalyzed a trend where researchers, proposing novel algorithms for the Virtual Machine (VM) consolidation problem, frequently adopt Beloglazov et al.'s algorithm as a baseline for comparison, showcasing enhancements through the lens of the PlanetLab experiment.

While CloudSim offers valuable features for energy consumption calculation and state modeling of data center components, challenges still exist. Achieving a high level of realism requires accurate input parameters, workload traces, and thorough validation against real-world data. Researchers should also be mindful of the limitations of simulation tools and the simplifications introduced to balance computational efficiency and accuracy. A more comprehensive simulation tool should allow for a detailed modeling of these transitions to better reflect the dynamic nature of data center operations. The representation of energy consumption from different states of PMs and VMs is crucial for accurate energy modeling. It can lead to inaccuracies in the simulation results. While CloudSim has been a cornerstone in cloud computing research, critiques regarding its I/O processing model [4], communication models [5], and the inaccuracy of power calculation [6] have prompted a reevaluation of simulation frameworks. Researchers, inspired by these critiques, have actively contributed to the development of more sophisticated tools that address the limitations identified in CloudSim.

SCORE (Simulator for Cloud Optimization of Resources and Energy Consumption) [7] is a cloud computing simulator designed to simulate cloud environments with a specific focus on optimizing the utilization of resources and minimizing energy consumption. While SCORE aims to provide a realistic simulation of data centers by offering simulation of energy-efficient monolithic and parallel-scheduling models and for the execution of heterogeneous, realistic and synthetic workloads, it doesn't provide a detailed networking models. Also, SCORE doesn't support VM migration policies which is important for modelling real life scenarios in cloud Systems. In real-world data centers, networking and VM migration play a crucial role in overall performance, and detailed simulation of network-related aspects is limited in SCORE.

DISSECT-CF [8] (DIScrete event baSed Energy Consumption simulaTor for Clouds and Federations) stands as a powerful simulation framework providing the ability to model energy consumption realistically. The simulator integrates sophisticated energy consumption models that consider the intricacies of the underlying infrastructure, including servers, networking equipment, and storage components. This allows researchers to gain insights into the energy usage patterns of Cloud Foundry environments. DISSECT-CF produces highly accurate simulation results in terms of finishing time and energy consumption. The reported error of just around 1% in most cases indicates a high level of precision in capturing the behavior of Cloud systems [6]. The ability of DISSECT-CF to deliver results much faster than CloudSim enhances its practical utility for researchers dealing with extensive simulations.

Based on the above discussion, we decided to use the DISSECT-CF simulator to produce a more realistic simulation of the planetlab workload trace. Table 1 shows a coparison of the cloudSimulators based on some features.

Table 1.
Features comparison of cloud simulators

	Features	CloudSim	SCORE	DISSECT-CF
1	**Communication model**	Limited	No	Full
2	**Network model**	Yes	No	Yes
3	**PM/VM model**	Yes	Only PM	Yes
4	**Resource scheduling model**	CPU and Network	CPU, Network, and Storage	CPU, Network, and Storage
5	**Migration policy**	Yes	No	Yes
6	**Distributed architecture**	No	Yes	No

In our previous work [9], we have implemented a new trace loading mechanism in DISSECT-CF simulator in order to load the planetlab workload trace and generate jobs that will be executed by the VMs based on their arrival time. The remainder of this work is structured as follows: Next section provides an overview of PlanetLab experiments in CloudSim including our observation on - to our knowledge - unrealistic behaviour. In later section, we elaborate on our experimentation and re-implementation process aimed at enhancing the realism of simulating PlanetLab workload traces and also presenting the results obtained concerning energy consumption and host utilization. Last section concludes the paper.

2. PLANETLAB WORKLOAD TRACE

PlanetLab serves as a global initiative dedicated to advancing research in distributed computing, it operates multiple co-located yet isolated user tasks, referred to as slices, through virtualization [2]. The CoMon monitoring system, developed to collect and process data from PlanetLab nodes [10], gathered valuable information, including the temporal evolution of slices' resource requirements, from approximately 600 nodes over several years. The only available remnants are the CPU usage data spanning 10 days and approximately 1000 virtual machines (VMs), preserved in the "*examples/workload/planetlab*" folder of CloudSim. Beloglazov et al. focused their experiments on CPU load, resulting in the retention of solely CPU load data. These data are stored in uncomplicated text files, each representing a single VM for a day. Each line within the file corresponds to a distinct measurement, indicating the CPU load as a percentage of the requested capacity. Typically, the CPU load was recorded at 5-minute intervals, leading to 288 measurements daily and, consequently, 288 lines per file. Table 2 shows the parameters used by Beloglazov et al. for evaluating their experiment.

Table 2.
parameters of the PlanetLab experiment, as defined by Beloglazov et al.

VM types	4
number of CPUs for each VM type	1, 1, 1, 1
CPU sizes [MIPS] for VMs	2500, 2000, 1000, 500
RAM sizes [MB] for VMs	870, 1740, 1740, 613
PM types	2
Number of CPUs for each PM type	2, 2
PM CPU capacities [MIPS]	1860, 2660
RAM sizes [MB] for PMs	4096, 4096
Power models for PMs	HP ProLiant ML110 G4, HP ProLiant ML110 G5

The VMs are initially created with requested resources adhering to the nominal values specified in Table 2. Consequently, these values govern the initial allocation of VMs to PMs. However, following the initial re-optimization phase, CPU values undergo modification in alignment with the PlanetLab trace, resulting in considerably reduced CPU loads. Additionally, the RAM utilization values for VMs are set to zero during this phase. This implies that, from the first re-optimization onward, VMs occupy significantly fewer CPU and RAM resources compared to their initial placement. Table 3 shows the power consumption characteristics of G5 and G4 servers for different CPU load intervals. For the planetLab experiment in CloudSim, The inclination of the power efficiency curve for G5 servers demonstrates a more favorable trend between 0% and 30%, compared to the range between 30% and 60%. This observed bias may result in suboptimal decisions by the algorithm [6]. To illustrate, in a scenario where the algorithm must choose between two G5 servers one with a CPU load of 20% and the other with a CPU load of 60%—for a VM with a CPU load corresponding to 10% of the PMs' capacity, the algorithm might favor the PM with a CPU load of 20% due to its perceived superior power efficiency. However, this contradicts the principle of VM consolidation, which advocates for consolidating as much load as possible onto already well-utilized PMs while attempting to free up lightly loaded PMs.

Table 3.
Power consumption (kWh) characteristic of PM types
for different CPU load intervals

PM type	**idle**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
G5	93.7	97	101	105	110	116	121	125	129	133	135
G4	86	89.4	92.6	96	99.5	102	106	108	112	114	117

In DISSECT-CF, the implementation of energy metering differs significantly. Unlike CloudSim, there is no straightforward method for querying the current power consumption of a Physical Machine (PM). Instead, each PM may be associated with an energy meter object that consistently monitors its energy usage. The energy consumption of the PM within two measurement points is computed based on the CPU processing activity during this interval, correlating with the average power consumption. Notably, in contrast to CloudSim, the frequency of power measurements can be independently configured from the re-optimization interval and can even vary between different energy meter objects. The selection of the power measurement frequency represents a crucial trade-off, as higher frequencies yield greater precision but also result in an increased execution time for the simulator.

3. EXPERIMENTAL RESULT

To reproduce the previously observed behavior, we opted to replicate the implementation of the PlanetLab experiment in DISSECT-CF. The goal was to facilitate a realistic simulation and compare the performance of the two simulators.

In our privious work, we introduced a new trace loading mechanism in the DISSECT-CF simulator. This mechanism aims to generate authentic jobs based on the PlanetLab workload trace, enabling these jobs to utilize the virtual machines (VMs) in the datacenter throughout the simulation process.

Furthermore, we incorporated two power models for G4 and G5 server types within DISSECT-CF. These models are designed to reflect energy consumption based on various parameters, including CPU load, network bandwidth, and memory utilization, as outlined in Table 3. Additionally, we implemented power transition generators to capture energy consumption by considering multiple states of both Physical Machines (PMs) and VMs.

Concerning the initial allocation of VMs to PMs, a VM allocation policy was implemented to mirror the behavior of CloudSim's VM allocation policy. This policy takes into account more realistic measurements, including RAM and bandwidth utilization of the PMs, in addition to the current CPU utilization.

For the sake of a straightforward comparison between the two simulators, we focused solely on the initial VM allocations without involving any VM migrations. The primary objective was to assess differences in energy consumption rather than delve into the decision-making processes related to VM migration by the VM scheduler.

Figure 1 depicts a comparison of energy consumption between CloudSim and DISSECT-CF utilizing an identical infrastructure setup. The increased energy consumption in DISSECT-CF is attributed to the consideration of additional factors that impact energy measurements in real-life scenarios. These factors encompass the energy consumed during state transitions of the Physical Machines (PMs) and the inclusion of other components such as RAM and network bandwidth in the VM-to-PM allocation process. It is noteworthy that to ensure a fair comparison between the two simulators, we configured the VM allocation so that

both have an identical VM-to-PM allocation map. Furthermore, DISSECT-CF exhibited significantly faster simulation times compared to CloudSim.

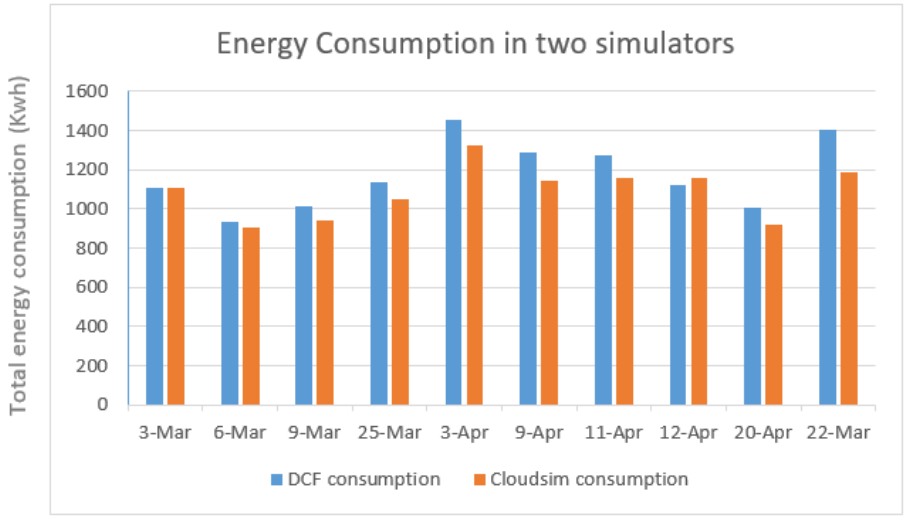


Fig. 1.
Energy consumption differences between
CloudSim and DISSECT-CF simulators

4. CONCLUSION

The ubiquity and intricacy of distributed computing systems underscore the significance of a simulation toolkit for developing, configuring, and assessing system performance. Given the challenges in accurately predicting power consumption for large systems, it is imperative for cloud simulators to address this complexity. This paper discussed our endeavor to port the PlanetLab experiment and its evaluation setup from one cloud simulator, CloudSim, to another, namely DISSECT-CF, with the aim of achieving a more realistic simulation. Our empirical findings affirm that DISSECT-CF holds a significant advantage in terms of execution time compared to CloudSim. Furthermore, DISSECT-CF offers a more comprehensive representation of energy-consuming aspects in the simulation. Our ongoing efforts involve the implementation of additional VM allocation and selection policies in DISSECT-CF, with the ultimate goal of achieving a complete implementation of the PlanetLab experiment.

REFERENCES

[1] Anton BELOGLAZOV and Rajkumar BUYYA: **Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers.** Concurrency and Computation: Practice and Experience, 24(13):1397–1420, 2012.

[2] Larry PETERSON, Andy BAVIER, Marc E. FIUCZYNSKI, and Steve MUIR: **Experiences building PlanetLab**. In Proceedings of the 7th Sym-

posium on Operating Systems Design and Implementation, pages 351–366, 2006

[3] Calheiros, RODRIGO N., et al.: **CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms.** Software: Practice and experience 41.1 (2011): 23-50.

[4] Hamza OUARNOUGHI, Jalil BOUKHOBZA, Frank SINGHOFF, Stéphane RUBINI and Erwann KASSIS: **Considering I/O processing in CloudSim for performance and energy evaluation.** In International Conference on High Performance Computing, pages 591–603, 2016.

[5] Alberto NÚÑEZ, Jose L. VÁZQUEZ-POLETTI, Agustin C. CAMINERO, Gabriel G. CASTAÑÉ, Jesus CARRETERO and Ignacio M. LLORENTE: **iCanCloud: A flexible and scalable cloud infrastructure simulator.** Journal of Grid Computing, 10(1):185–209, 2012.

[6] MANN, Zoltán Ádám: **Cloud simulators in the implementation and evaluation of virtual machine placement algorithms.** Software: Practice and Experience 48.7 (2018): 1368-1389.

[7] Fernández-Cerero, DAMIÁN, et al.: **SCORE: Simulator for cloud optimization of resources and energy consumption.** Simulation Modelling Practice and Theory 82 (2018): 160-173.

[8] Gabor KECSKEMETI: **DISSECT-CF: A simulator to foster energy-aware scheduling in infrastructure clouds.** Simulation Modelling Practice and Theory, 58:188-218, 2015.

[9] RJEIB, Hasanein, and Gábor KECSKEMÉTI: **An investigation on implementing a scenario on different cloud simulators.** MULTIDISZCIPLINÁRIS TUDOMÁNYOK: A MISKOLCI EGYETEM KÖZLEMÉNYE 12.3 (2022): 256-263.

[10] KyoungSoo PARK and Vivek S. PAI: **CoMon: a mostly-scalable monitoring system for PlanetLab**. ACM SIGOPS Operating Systems Review, 40(1):65-74, 2006.