View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Birkbeck Institutional Research Online

Soft Computing manuscript No.
(will be inserted by the editor)

Analysis of Power Consumption in Heterogeneous Virtual Machine Environments

Catalin Negru · Mariana Mocanu · Valentin Cristea · Stelios Sotiriadis · Nik Bessis

Received: date / Accepted: date

Abstract Reduction of energy consumption in Cloud computing datacenters today is a hot a research topic, as these consume large amounts of energy. Furthermore, most of the energy is used inefficiently because of the improper usage of computational resources such as CPU, storage, and network. A good balance between the computing resources and performed workload is mandatory. In the context of data-intensive applications, a significant portion of energy is consumed just to keep alive virtual machines or to move data around without performing useful computation. Moreover, heterogeneity of resources increases the degree of difficulty, when try to achieve energy efficiency. Power consumption optimization requires identification of those inefficiencies in the underlying system and applications. Based on the relation between server load and energy consumption, we study the efficiency of data-intensive applications, and the penalties, in terms of power consumption, that are introduced by different degrees of heterogeneity of the virtual machines characteristics in a cluster.

Keywords Cloud computing · Energy-Efficiency · Virtualization · Data Intensive-Applications

Catalin Negru - Principal Author,
Computer Science and Engineering Department, University "Politehnica" of Bucharest,Romania E-mail: catalin.negru@cs.pub.ro

Mariana Mocanu
Computer Science and Engineering Department, University "Politehnica" of Bucharest, Romania E-mail: mariana.mocanu@cs.pub.ro

Valentin Cristea
Computer Science and Engineering Department, University "Politehnica" of Bucharest, Romania E-mail: valentin.cristea@cs.pub.ro

Stelios Sotiriadis - Corresponding Author,
The Edward Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada E-mail: s.sotiriadis@utoronto.ca

Nik Bessis
Department of Computing, Edge Hill University, UK E-mail: nik.bessis@edgehill.ac.uk

1 Introduction

Nowadays, Cloud computing datacenters consume large amounts of energy. By 2020 these will use approximately 140 kilowatt-hours. The usage on large scale of the services provided by Cloud led to this situation. So, there is an urgent need to reduce the consume of energy.

A possible way to reduce the energy consumption is to improve resource utilization. Virtualization technology, permits the independence of servers and offers a new way to improve the data center energy efficiency by assignment of multiple virtual machines (VMs) to a single physical server. The problem is represented by the low utilization of virtual machines. Resources, such as CPU, memory, storage, and network, consume energy even when they are in idle state [2].

Resource provisioning problem is very challenging as we encounter a great diversity of workloads (e.g. computationally-intensive, data-intensive, and hybrid), different usage patterns (e.g. static, periodic, once-in-a-lifetime, unpredictable, and continuously changing), and virtual machine heterogeneity. For instance in the case of data-intensive applications a significant portion of energy is used just to keep virtual machines alive or move data around without performing a useful computation. Furthermore, reducing power consumption at the data center level has serious implications over the usage cost.

Few important questions arise when talking about power efficiency related to workload type and usage patterns. What happens when we have to deal with workloads that are computational intensive or data-intensive?How is better regarding the energy consumption, to use virtual machine instances, with low resources (e.g. CPU, RAM, Disk and bandwidth) and to process tasks in a longer time, or to use virtual machine instances with high resources and to process tasks in a shorter time? Also, what are the implications of virtual machine heterogeneity over the energy consumption? So, in order to study the optimization of power consumption, we need in the first place, to identify the inefficiencies in the underlying system. So, heterogeneity of virtual machines has a great deal of importance in power consumption.

Applications for smart-cities and cyber-infrastructures are data intensive and are I/O bounded. These dedicate a significant part of execution time to the data movement process, and in consequence need a high bandwidth data access rate. If the available bandwidth is less than required, the CPU is held idle until data sets are available. So, a virtual machine that is idle for a certain amount of time, will consume much more energy. Additionally, a heterogeneous environment contribute to the waste of energy especially in the case of data-intensive applications. This happens because of the resource characteristics diversity such as CPU, bandwidth and RAM memory in virtual machines.

Also, Big Data applications can help to uncover the fine interactions between data. In this way we are allowed to manipulate hidden, often-counterintuitive, levels that directly impact different domains and activities. Moreover, these applications can bring up new opportunities for business and consumer through modern marketing and networking technologies using an inclusive social and technological environment [35], [37]. Also, there are tools that supports managers in identifying forthcoming disruptive technologies and provide them with tailored strategic options [28].

Fields such as environmental research, disaster management and information in relief operations, decision support systems, crowd-sourcing, citizen sensing and sensor web technologies, need to make use of new and innovative tools and methods for Big Data, in order to be more efficient. So, we must be able to analyze all data in order to get the promoted benefits. For instance, a decision support system can give better and accurate indications in a crisis situation.

As shown by the authors of [53] the measurement of energy consumption in a virtual machine can be made by measuring the usage of the machine. So, there is a direct relation between server load and consumption of energy. Based on this relation we try to evaluate what are the penalties in terms of power consumption introduced by the heterogeneity of virtual machines.

This paper extends the work presented in [39]. The main contributions can be summarized as follows:

- we analyze the methods for power metering at virtual machine level;
- we identify inefficiencies of big-data applications;
- based on the relation between virtual machine load and power consumption, we present an approach for evaluating the impact of virtual machine heterogeneity on power consumption in a datacenter;

The paper is structured as the following: In Section 1, is presented the importance of power consumption optimization in Cloud computing data centers. Also, here are presented the questions and implications that arise when performing power consumption optimization. In Section 2, are presented previous related work for the power consumption optimization of virtual machines. Section 3 presents our approach for the impact evaluation of the virtual machine heterogeneity over the power consumption in a Cloud computing datacenter. Section 4 presents the Big Data applications and key issues regarding the energy consumption. In Sect. 5, the experimental setup, use case scenarios, results and analysis of the obtained results are presented. Finally in Sect. 6, the conclusions and future work are presented.

2 Related work

The problem of virtual machines efficient power consumption in Cloud computing infrastructures is very intense studied. Given the heterogeneous nature of resources, workloads and, usage patterns this is still a challenging problem. Surveying the literature we can distinguish few important research directions for power efficient Cloud computing.

One research direction refers to the methods and technologies for operation efficiency at the hardware level, meaning computer and network infrastructure. Technologies, such as SpeedStep [20], PowerNow [11], Cool'n Quiet [52] or DemandBased Switching [41] has been developed. Also, techniques like dynamic voltage scaling [43] have been applied in different provisioning and scheduling algorithms and workload consolidation techniques to minimize the power consumption. Moreover, frameworks for reducing power consumption in computer networks and network-wide optimization algorithms have been proposed. In [40] is proposed a two-level control framework providing local control mechanisms that are implemented at the network device level and network-wide control strategies implemented at the central level.

Another research direction that we have identified in the scientific literature refers to virtual machine placement problem. Also, different methods and algorithms has being proposed. For instance, in [19] the authors propose an algorithm that use methods like dynamic programming and local search in order to place in an energy-efficient way, previous created copies of virtual machines in order to meet the QoS, on the physical servers. An algorithm for virtual machine placement, designed to increase environmental sustainability in the context of distributed data centers with different carbon footprint and power utilization efficiency is presented in [23]. The results obtained from the simulation shows a reduction of CO_2 and power consumption, maintaining the same level of quality of service. Furthermore, a multi-objective ant colony system algorithm for minimization of total resource wastage and power consumption is proposed in [17]. The authors compare the algorithm with existing multi-objective genetic algorithm and two single-objective algorithms. In [21] is proposed a decentralized strategy for genetic scheduling in heterogeneous environments that uses a combination of genetic algorithms and lookup services for obtaining a scalable and highly reliable optimization tool.

Energy-efficient scheduling algorithms that assign virtual machines to physical machines represent another research direction that we have identified. For example, an algorithm, that aim to minimize total power consumption of physical machines in the datacenter, by assigning efficiently virtual machines is presented in [47]. The results obtained show 24.9% power saving and nearly 1.2% performance degradation. Also, an algorithm called Dynamic Round-Robin for energy-aware virtual machine scheduling and consolidation is proposed in [31]. Compared with other strategies such as Greedy, Round-Robin and PowerSave implemented in Eucalyptus, this reduces a significant amount of power. In [33] the authors propose and implement a virtual machine scheduling heuristics that take into consideration load-balancing and temperature-balancing with the aim of reducing the energy consumption in a Cloud data center.

Energy efficient, data-aware scheduling is also a major research direction. In Cloud computing, it pose additionally challenges, as data is stored and accessed on a large scale from distributed servers. In this situation the energy consumption reduction represents the principal scheduling objective. In [26], the authors deal with the problem of independent batch scheduling in grid environment as a bi-objective minimization problem with makespan and energy consumption as the scheduling criteria. Also in [27],are presented two implementations of classical genetic-based data-aware schedulers of independent tasks submitted to the grid environment. In [5] is optimized the energy efficiency of message exchanging for service distribution in interoperable infrastructures. The authors consider two use cases. First, a requester sends messages to all interconnected nodes and gets messages only from resources available to execute it and second, the requester sends one message for all of the jobs of its local pool and gets a respond from available nodes, and then obtainable resources are ranked and hierarchically categorized based on the performance criterion e.g. latency competency. Also in [6] the authors propose a novel message-exchanging optimization model to reduce energy consumption in distributed systems. They aim to achieve the optimization of the energy consumption for communication and to improve the overall system performance. The intercloud concept [49] encompasses the interconnectivity of node and authors of [5], [6] developed the inter-cloud meta-scheduling simulation framework [48] to evaluate the energy efficiency of message exchanging.

Workload consolidation represents also an important research direction, as it permits to place the workload on fewer physical machines, taking into consideration as a principal parameter the machine load. In this way is achieved the reduction of the power consumption. Usually, the workload placement problem is modeled as a multi-dimensional bin-packaging problem, as expressed in [45], [42], [29]. Moreover, meta-heuristics such as Ant Colony Optimization [14], [15], [17], Genetic Algorithms [25], [46], [44], [24] are used for power consumption optimization.

More recent research directions are in the domain of security. One example is prevention of some energy oriented distributed denial of service (e-DDoS) attacks. These attacks are characterized by the fact that do not produce direct damage or block the activity of the targeted infrastructure, but instead generate an anomalous and sustained power consumption on the target side. In this way, IT equipment and facilities such as air conditioning, heating and ventilation are affected, their lifetime is reduced significantly. Also, this type of attacks increases energy bills, drastically [16].

The previous related works do not take specifically into consideration and do not evaluate the heterogeneity degree of virtual machines when proposed different resource management methods such as resource allocation, job scheduling and workload consolidation techniques. Quantifying the penalties that are introduced by the different degrees of heterogeneity can be used further as the input parameter in scheduling and provisioning algorithms for energy consumption optimization. Based on previous related works we build a simple taxonomy presented in Fig. 1.

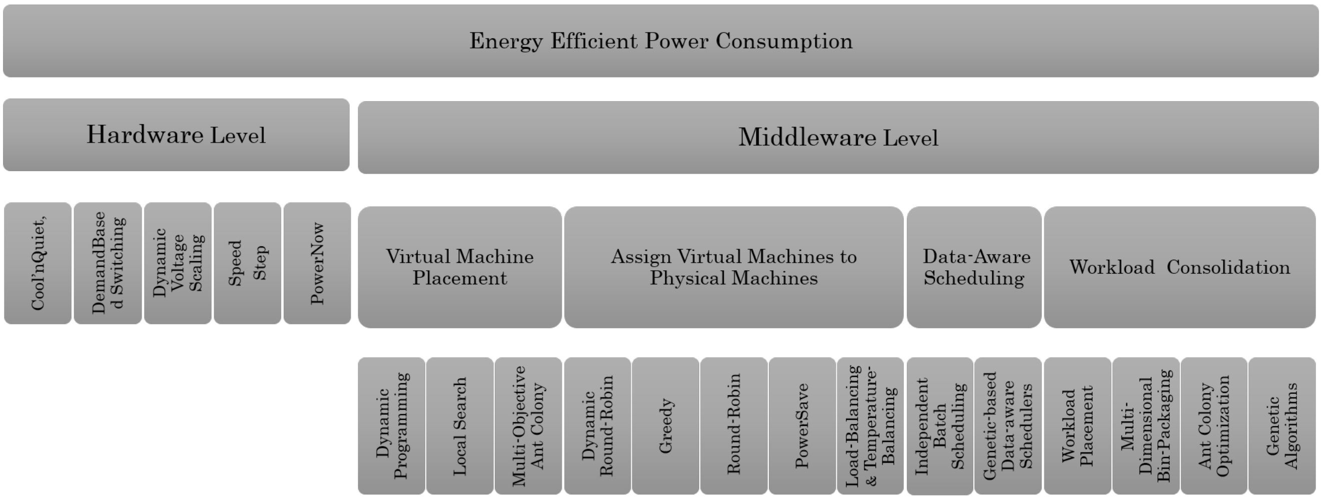

Fig. 1 Energy efficient power consumption taxonomy

3 Virtual machine power metering

Server virtualization represents one step further in power consumption optimization in Cloud computing datacenters, permitting effective, efficient energy management. In order to optimize power consumption, we have to measure the power consumption on per-virtual machine basis in an accurate and efficient way. In [36] the authors provide a comprehensive survey of proposed approaches for estimating the power consumption of single-core as well as multi-core processors, virtual machines, and an entire server.

The virtual machine power models for power consumption metering proposed in literature can be classified in two categories: utilization-based models [32],[8], [3] and performance-monitor-counter models [30], [7], [4]. The first category of models assume that a server resources (e.g. CPU, memory, disk) consume of energy is linear with his utilization [36]:

$$P_{server} = P_{static} + \sum_{j \in J} (k_j \cdot U_j), \tag{1}$$

where:

- P_{static} - fixed power consumption when there is no workload;
- U_j - utilization of physical component;
- k_j - the dynamic power coefficient;
- $J = CPU, RAM, Disk, I/O$ - set of power consuming components.

Starting from equation 1, can be obtained the most used virtual machine power model:

$$P_i^{vm} = \frac{P_{static}}{M} \cdot \sum_{j \in J} (k_j \cdot U_j), \tag{2}$$

where:

- W_i - the processor utilization of the virtual machine;
- M - the number of active VMs on a server.

Performance monitor counter models are based on software components called counters, that monitor the performance of the physical server offering a real-time method for power consumption monitoring. These counters are supported by all modern processor architectures. The power model for a virtual machine using performance monitor counter models can be expressed as follows:

$$P_i^{vm}(t_1, t_2) = \sum_{j \in J} P_{ij}^{vm}(t_1, t_2), \tag{3}$$

where:

- $P_{ij}^{vm}(t_1, t_2)$ - the power consumption consumed by physical component j in time interval $[t_1, t_2]$.

The authors of [1], in order to formulate the problem of power consumption minimization, propose the following objective function:

$$P(\pi) = \sum_{i \in [1,m]: A_i \neq \emptyset} (\mu \sum_{d_j \in A_i} (l(d_j)^{\alpha} + b), \tag{4}$$

where:

- $\pi = \{A_1, \dots A_m\}$ - set of virtual machines;
- $l(d_j)$ - load of a virtual machine;
- μ - dynamic power coefficient;
- b - static power consumption.

Then the power consumption function for a set of virtual machines can be express as following:

$$P(\pi) = \sum_{i=1}^{m} f(l(A_i)) \tag{5}$$

Further in our paper we used Equation(4) to quantify the power consumption for our considered set of virtual machines.

4 Big Data Applications

Big Data represents a paradigm for the advancing trends in technology that propose a new approach in the process of decision-making. Moreover, it means to deal with huge amounts of heterogeneous data, and ask for a pipeline of processing operations. The aim is to offer support in the decision-making process. An important challenge, besides the large volumes and high-speed production rates (e.g. velocity), is raised by the data heterogeneity (e.g. variety). The most important Big Data characteristics are volume, variety, velocity, value, veracity, volatility and vicissitude.

Data volume is the main challenge, because traditional storage systems(e.g. relational databases) did not succeed to handle volumes of data in terms of terabyte and petabyte levels. Developments in different areas such as Cyber-infrastructures, Smart Cities, e-Health, Social Media, Web 3.0, etc., has led to large amounts of data, and these data are often unstructured or semi-structured, with a high level of heterogeneity.

Variety characteristic refers to different data formats and sources such as data from sensors, documents, emails, social media texts, mobile devices etc. Velocity refers to the data acquisition rate, as data can be acquired at different speeds. The value property illustrates the potential gain of data, obtained after some processing operations.

Veracity characteristic describes how accurate is data collected from different sources. For example, data gathered from a social media website have a specific degree of accuracy, in other words represents the uncertainty in data. Veracity ensure that used data are trusted, authentic and protected for unauthorized access and modification. Furthermore, the data must be secured during the whole their lifecycle from collection to storage [13].

Vicissitude property refers to the challenge of scaling Big Data complex workflows. This property signifies a combination between the large volume of data and the complexity of processing workflow, which prevent to gather useful insights in data [18].

4.1 Real-Time Data Access and Processing

The batch data processing has the following workflow of operations: data collection, send to the processing system, process, and extract the results. This is an efficient way to process high volumes of data. This type of processing requires separate applications for input, process and output.

In contrast with batch data processing, real time data processing implies a input, process and output of data in a continuous mode. This means that data need to be processed in a small time. Moreover, there are two categories of data:

– predicted data - that has value at the given moment in time);
– measured data - value remains forever (e.g. sensor data) which represents historical data.

Mining the instantaneously valued data requires a real time platform. Also, a method of dynamic pattern identification for logically clustering log data is needed. The method must be a real time and generalized solution to the process of log file management and analysis. For instance, a monitoring platform for water data management needs to access distributed data sources (e. g. sensor networks, mobile systems, data repository, social web, and so on). Next, this data has to be processed for preventing natural disasters such as water pollution and to alert the possible affected people [38].

Frameworks such as MapReduce [12], Dryad [22] are used for large-scale data processing. Users write parallel computations with the aid of high-level operators, without paying attention to data distribution or fault tolerance. These systems are batch-processing systems and are not designed for real-time processing and this is an important disadvantage. A possible solution for real-time data streaming processing is provided by Storm [51] and Spark [50]. Storm is used at Twitter for real-time distributed processing of stream data.

In the context of Big Data traditional technologies for the data processing and analysis are not efficient. There is an imperative need to discover valuable knowledge in data in order to gain help in decision making process. Different challenges are encountered when dealing with big data handling, from data capture to visualization. Moreover, sources of data are heterogeneous, geographically distributed, and unreliable, being susceptible to errors. As a consequence we face with databases that are populated with inconsistent, incomplete and noisy data. Therefore, several data preprocessing techniques, such as data reduction, cleaning, integration, data transformation, must be applied to remove noise and correct inconsistencies in order to help in decision-making process.

NoSQL databases systems highlights a series of advantages for data handling compared with relational database systems. First advantage is represented by the fact that storage and management parts are independent one from another. The data storage part, (e.g.

key-value storage), focus on the scalability and high-performance, while the management part, is built on low-level access mechanisms, meaning that related tasks to data management can be implemented in the application layer. For instance, three operators are defined for data cleaning tasks:

– fuzzy lookup - used to perform record matching;
– fuzzy grouping - used for deduplication;
– column segmentation - uses regular expressions to segment input strings.

In this way, NoSQL databases provide flexible mechanisms for data handling and application developments. Moreover, deployments can be easily updated.

Many fields(e.g. scientific, business, etc.) can benefits from Big Data with the condition to solve the challenges, which arise in data storage, capture, cleaning, analysis and visualization.

4.2 Energy Efficient Processing in Big Data Platforms

A Big Data processing platform is composed by two parts, a job manager and a storage manager. First one coordinates the processing nodes, and the second one coordinate the storage nodes. An examples is Apache Hadoop [10] composed by a set of open source applications that are used together to provide a Big Data processing solution. HDFS and YARN are the main component of Hadoop. YARN coordinates the processing nodes and HDFS the storage nodes. Fig. 2 presents the general architecture of Big Data platforms.

Hadoop distributed file system (HDFS) is organized in clusters where each cluster consists of a name node and several storage nodes. A large file is split into blocks and name node takes care of the persisting parts on data nodes. The name node maintains metadata about the files and commits updates to a file from a temporary cache to the permanent data node. The data node does not have knowledge about the full logical HDFS file; it handles locally each block as a separate file. Fault tolerance is achieved through replication; optimizing the communication by considering the location of the data nodes (the ones located on the same rack are preferred). A high degree of reliability is realized using "heartbeat" technique (for monitoring), snapshots, metadata replication, checksums (for data integrity), and re-balancing (for per-formance) [9].

YARN, the version 2.0 of MapReduce, implements a master/slave execution of processes with a JobTracker master node and a pool of Task-Trackers that do the work. JobTracker has two main responsibilities, management of resources and job scheduling/monitoring.

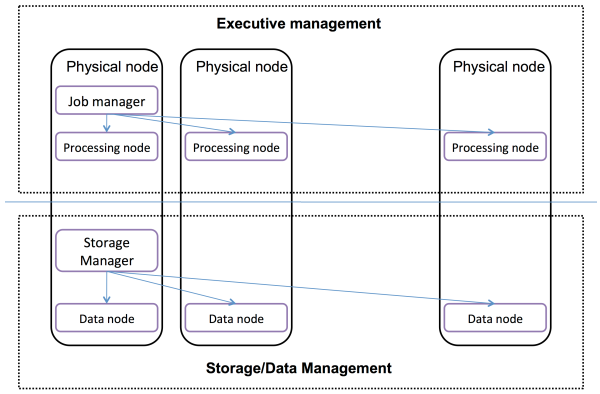

Fig. 2 Organization of resources in a big-data platform

Also, there is a global resource manager (RM) per application, and an Application Master (AM). The slave has a node entity named Node Manager (NM) which is doing the computations. The AM negotiates with the RM for resources and monitors task progress. Other components are added on top of Hadoop in order to create a Big Data ecosystem capable of configuration management (Zookeeper), columnar organization (HBase), data warehouse querying (Hive), easier development of MapReduce programs (Pig), and machine learning algorithms (Mahout).

In Big Data platforms each layer (e.g. operating systems, databases and environments, computing solutions, data operations and analytics and data sources) of the processing stack impose power issues, as can see in the Fig. 3. First, the usage of virtual machine components (e.g. CPU, storage, RAM and network) have a direct influence on power consumption of the hardware components. Operating systems, databases and environments have direct impact on the hardware usage and thus on power consumption. Computing solutions produce the system behavior power consumption factor. Data operations and analytics have effect on data movement and processing. Finally, data sources produce effects in memory and storage components.

So, in order to optimize the energy consumption of the underling hardware, every processing step should be optimized. For instance, this could mean optimized reads and writes from memory and storage, efficient data movement and processing, improved system behavior.

5 Impact Evaluation on Power Consumption of Virtual Machines Heterogeneity

In the first place, in order to evaluate the impact of virtual machine heterogeneity on power consumption at the datacenter level we performed several experiments

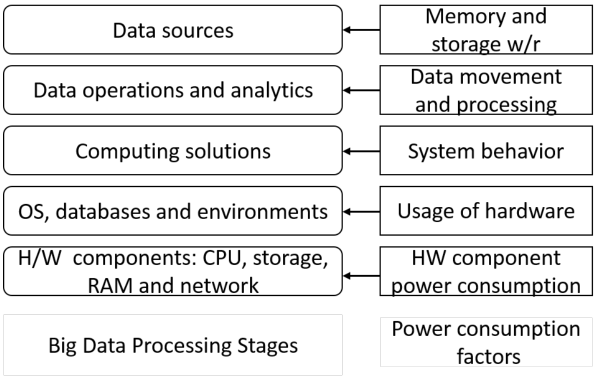

Fig. 3 Energy efficiency issues Big Data processing

with different degrees of heterogeneity. To achieve our goal, we used multiple sets of virtual machines having different heterogeneity degrees. Each set is composed by four instances. Further, we increase gradually the degree of heterogeneity and calculated the power consumption for different degrees of heterogeneity. Regarding the workload type, we chose to perform a data-intensive job, all machines sending and receiving simultaneously a file of $200MB$. At every 15 seconds we get the server load on every machines until the job is done.

In the first experiment we start four identical machines and performed the data transfer job. In the second experiment we used three identical instances and the forth instance was different. For the third experiment we used two machines of one type and two of other type. In the fourth experiment we used three different sets of virtual machines. In the last experiment, all machines were different, having the highest degree of heterogeneity.

In the second place, we want to evaluate the difference in terms of power consumption between two different Hadoop clusters with same computing power when running the same workload. Regarding the workload type we run three benchmarks, TeraGen, TeraSort and TeraValidate. TeraGen was used to generate 10GB of random distributed data, TeraSort was used to sort the data and TeraValidate was used to validate the sorted data. These benchmarks are commonly used to measure MapReduce performance of an Apache Hadoop cluster.

We used the following formula in order to calculate the power consumption and it is based on Equation 5:

$$P(4) = \sum_{i=1}^{4} (l(i)^3 + 0.1), \qquad (6)$$

where:

- i represents a virtual machine instance;
- $\alpha = 3$;
- $b = 0.1$;
- $\mu = 1$.

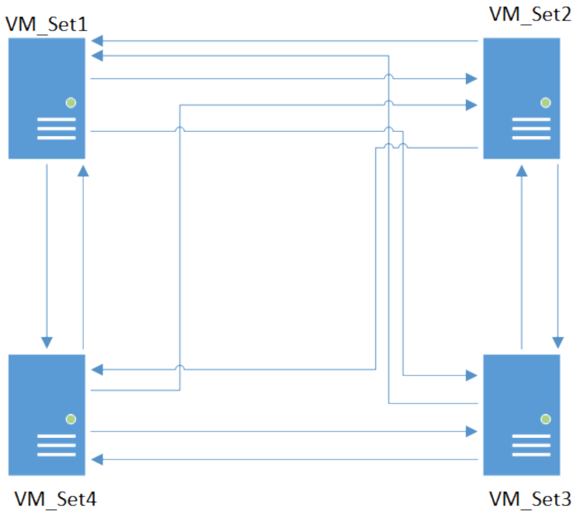

Fig. 4 Full mesh logical topology

5.1 Experimental Setup

We create two different experimental setups. First, in order to perform a data-intensive job, we interconnected all machines together in a full mesh logical topology as shown in Fig. 4. Further, each virtual machine sends and receive data. We used four types of virtual machines from Microsoft Azure Cloud [34] with different CPU and RAM memory characteristics presented in Table 5.1: Basic A0, Basic A1, Basic A2 and Basic A3 instance types.

Secondly, in order to evaluate power consumption in a Hadoop MapReduce cluster, we used two different Hadoop clusters with same computing power. Cluster 1 has six worker nodes (with the following characteristics:1 virtual CPU and 3.75 GB RAM memory) and one master node (with 2 virtual CPUs and 7GB RAM memory). Cluster 2 has three worker nodes and one master node. All virtual machines has the same configuration, 2 virtual CPUs and 7GB RAM memory.

In all experiments performed, we get the average load of the virtual machines using the "uptime" command. The average load Linux systems, represents a measurement of the computational work that the system is performing. Moreover, Linux systems counts processes waiting for other resources than CPU such as, processes that wait to read from or write to the disk.

VM type	Cores	RAM (GB)	Disk (GB)
Basic A0	0.25	0.75	30
Basic A1	1	1,75	30
Basic A2	2	3,00	30
Basic A3	4	7,00	30

Table 1 Azure instances types that have been used in experiments

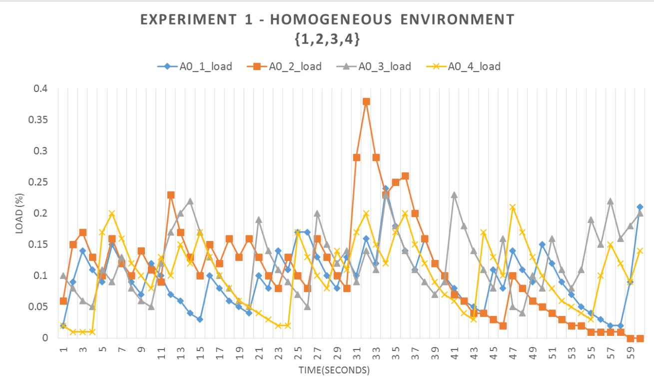

Fig. 5 Experiment 1: Homogeneous environment

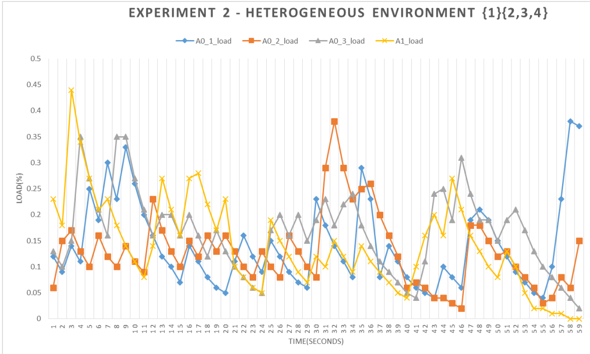

Fig. 6 Experiment 2: Heterogeneous environment

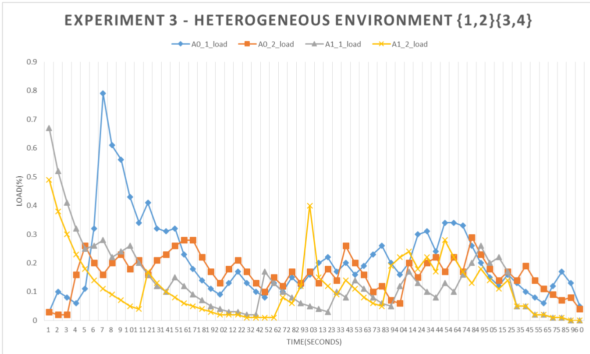

Fig. 7 Experiment 3: Heterogeneous environment

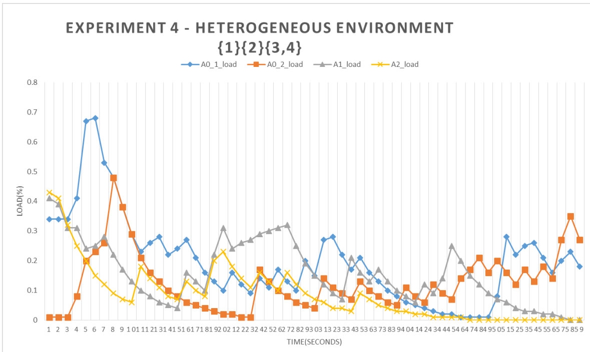

Fig. 8 Experiment 4: Heterogeneous environment

5.2 Results

5.2.1 Evolution of server load

In order to understand the effect of heterogeneity on power consumption of the machines, we present the evolution of server load in a for different degrees of heterogeneity. In the first experiment, we considered a homogeneous environment. In this setup the type of virtual machines is "Basic A0". The evolution of server load is presented in the Fig. 5. As can we see the load has same evolution pattern on all machines.Moreover, al machines finish the data transfer in the same time.

In the second experiment, we considered an heterogeneous environment with three identical sets of virtual machines formed from "Basic A0" instance types and one different with "Basic A1" instance type. The evolution of server load is presented in Fig. 6. We can observe a more random evolution pattern with large fluctuations of the load. "Basic A1" instance types finish first the transfer and have to wait for the rest of instances to finish the data transfer.

In the third experiment we increased the degree of heterogeneity. We interconnected two sets of virtual machines with different characteristics. First set has "Basic A0" instance types and the second one has "Basic A1". In the Fig 7 is presented the evolution of server load. As can we observe the sets with identical machines has the same evolution pattern. Moreover, the virtual machines with higher performance characteristics (e.g. Basic A1) finish the transfer before the other instance types. These have to wait for the slower machines to finish the transfer.

For the fourth experiment we have increased the degree of heterogeneity. We've interconnected three different sets of instances (e.g. Basic A0, A1 and A2). The evolution of server load is presented in Fig. 8. As can we see an irregular pattern of the load evolution and the range of load values increased. Also, "Basic A2" instance type finish first the transfer followed by the "Basic A1" and "Basic A0".

In the fifth experiment, we considered an environment with the highest degree of heterogeneity. In this setup we interconnected four different sets of virtual machines. In Fig. 9 is presented the evolution of server

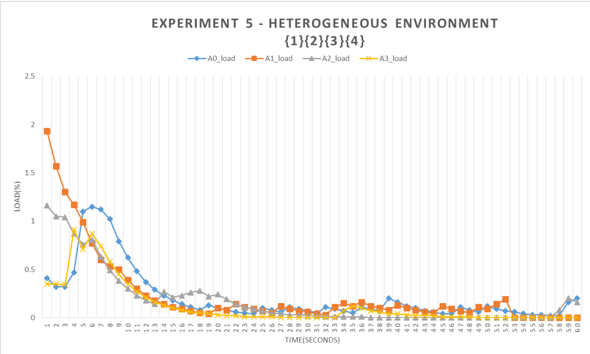

Fig. 9 Experiment 5: Heterogeneous environment

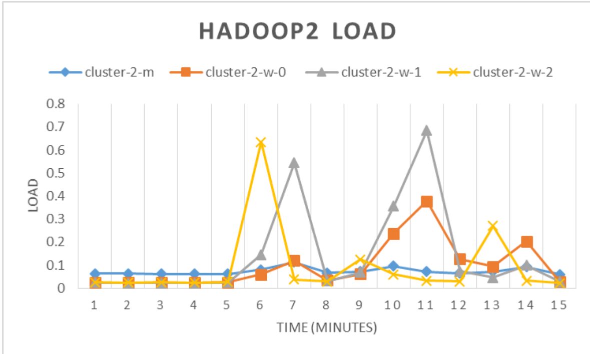

Fig. 11 Hadoop cluster 2 load evolution

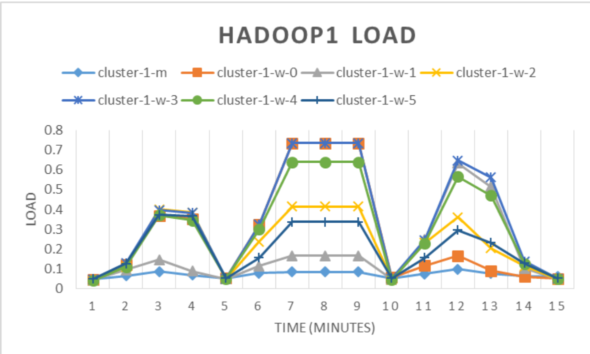

Fig. 10 Hadoop cluster 1 load evolution

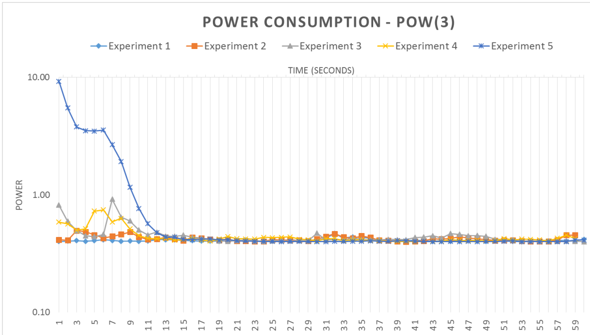

Fig. 12 Power consumption for each experiment

load in this type of environment. The powerful instances A2, and A3 finish the job much faster, approximatively after 100 second and have to wait for the other less powerful virtual machines.

In the second part of our experiment we measured the evolution of load on two different Hadoop clusters, while performing a MapReduce benchmark(e.g. TeraGen, TeraSort and TeraValidate.) The load evolution for the first cluster is presented in Fig. 10 and for the second cluster is presented in Fig. 11.

Comparing the load evolution of the two clusters we can observe that the first cluster need more time to generate (approximately five minutes more), sort and validate the data than the second cluster. We can also observe that the cluster nodes are not equally balanced.

5.2.2 Power Consumption

In the Fig. 12 is presented in an aggregated way the evolution of energy consumption over the execution of job. As can we see the environment with highest degree of heterogeneity (Experiment 5) consume more power than the homogeneous environment (Experiment 1). Furthermore, we observe that for the Experiment 5 in the first 15 seconds, which is the time that is needed for the powerful machines to finish the transfer, is consumed an important quantity of energy. In the remaining time these stay in an idle state while have to wait for the slower machines to finish the transfer and thus consume energy while do not make any useful work.

So, we can conclude that the heterogeneous environments consume more power because the virtual machines with higher resource characteristics finish the transfer much faster than slower machines and have to wait in an idle for the other virtual machines to finish the transfer, consuming power without performing useful computation. The key issue is to reduce the idle time of the virtual machines.

In the Fig. 13 is presented the power consumption for the each experiment performed. As can we see a fully heterogeneous environment consume approximately twice the power of a homogeneous environment.

The power consumption evolution of the load for the nodes of Hadoop cluster 1 is presented in Fig. 14 and for Hadoop cluster 2 is presented in Fig. 15. As can we see, for the first cluster the power consumption distribution on nodes is unbalanced. There are two nodes (e.g. cluster-1-w-3 and cluster-1-w-5) that consume the same amount of power. Comparing the power consumed by the two clusters we can observe that cluster 1 consume approximately twice the power of the cluster 2. So, we can conclude that in order to optimize power consumption in Big Data processing platforms we must

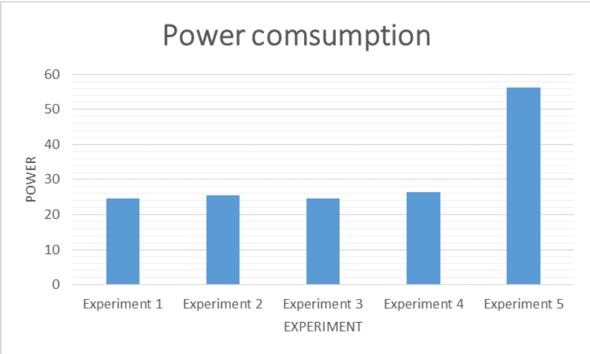

Fig. 13 Power consumption for each experiment

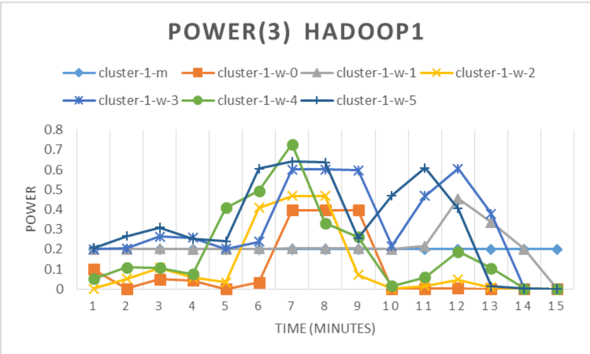

Fig. 14 Hadoop cluster 1 power consumption

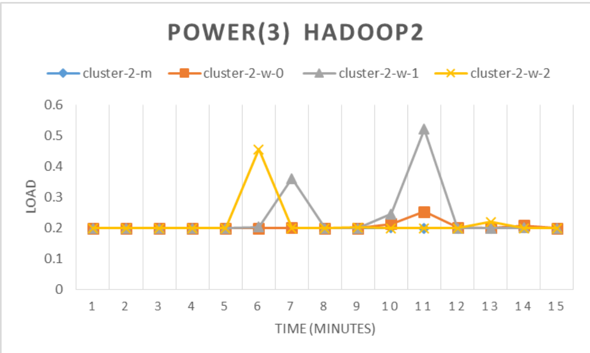

Fig. 15 Hadoop cluster 2 power consumption

find the best combination between the configuration of computing resources and workload type.

6 Conclusions and Future Work

In this paper we presented an approach for evaluation of the heterogeneity degree impact on power consumption for a set of virtual machines in a Cloud computing environment in the case of data-intensive jobs in virtual machines.

A good balance between workloads, usage patterns and virtual machine computing power is mandatory in order to achieve power efficiency. If the virtual machine utilization is low, and is still running, more power is consumed. As a consequence virtual machines should be dynamically adjusted to match the characteristics of the other virtual machines that performs the job. In this way the degree of heterogeneity decrease and the virtual machines can finish the data transfer in simultaneously, reducing the energy consumed. The key issue is to reduce the idle time for the used resources.

Results show that the power consumption is proportional with heterogeneity degree. This happens because of the fact that powerful virtual machines finish the transfer much more quickly that the less powerful ones, and wait just to receive the data.

As we showed through the paper heterogeneity degree has a big impact on power consumption for a set of virtual machines that perform data-intensive tasks. This has also impact on cost, because the cost of energy represent an important component of the cost for services and resources.

For future work more type of workloads can be used in order to quantify more accurate the impact on power consumption. Furthermore, based on these results we can build a job scheduling algorithm that take into consideration the heterogeneity degree for the set of instances that must execute scheduled job, in order to optimize the power consumption at data center level.

Acknowledgements

The work has been funded by the projects: DataWay: Real-time Data Processing Platform for Smart Cities: Making sense of Big Data, PN-II-RU-TE-2014-4-2731; CyberWater grant of the Romanian National Authority for Scientific Research, CNDI-UEFISCDI, project number 47/2012;

Compliance with Ethical Standards

The authors of this paper Catalin Negru, Mariana Mocanu, Valentin Cristea, Stelios Sotiriadis and Nik Bessis declare that they have no conflict of interest.

This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Jordi Arjona Aroca, Antonio Fernández Anta, Miguel A Mosteiro, Christopher Thraves, and Lin Wang. Power-efficient assignment of virtual machines to physical machines. *Future Generation Computer Systems*, 2015.

2. Luiz André Barroso, Jimmy Clidaras, and Urs Hölzle. The datacenter as a computer: An introduction to the design of warehouse-scale machines. *Synthesis lectures on computer architecture*, 8(3):1–154, 2013.
3. Andreas Berl and Hermann De Meer. An energy consumption model for virtualized office environments. *Future Generation Computer Systems*, 27(8):1047–1055, 2011.
4. Ramon Bertran, Yolanda Becerra, David Carrera, Vicenç Beltran, Marc Gonzàlez, Xavier Martorell, Nacho Navarro, Jordi Torres, and Eduard Ayguadé. Energy accounting for shared virtualized environments under dvfs using pmc-based power models. *Future Generation Computer Systems*, 28(2):457–468, 2012.
5. Nik Bessis, Stelios Sotiriadis, Florin Pop, and Valentin Cristea. Optimizing the energy efficiency of message exchanging for service distribution in interoperable infrastructures. In *Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on*, pages 105–112. IEEE, 2012.
6. Nik Bessis, Stelios Sotiriadis, Florin Pop, and Valentin Cristea. Using a novel message-exchanging optimization (meo) model to reduce energy consumption in distributed systems. *Simulation Modelling Practice and Theory*, 39:104–120, 2013.
7. W Lloyd Bircher and Lizy K John. Complete system power estimation using processor performance events. *Computers, IEEE Transactions on*, 61(4):563–577, 2012.
8. Ata E Husain Bohra and Vipin Chaudhary. Vmeter: Power modelling for virtualized clouds. In *Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on*, pages 1–8. Ieee, 2010.
9. Dhruba Borthakur. Hdfs architecture guide. *HADOOP APACHE PROJECT http://hadoop. apache. org/common/docs/current/hdfs design. pdf*, 2008.
10. Dhruba Borthakur, Jonathan Gray, Joydeep Sen Sarma, Kannan Muthukkaruppan, Nicolas Spiegelberg, Hairong Kuang, Karthik Ranganathan, Dmytro Molkov, Aravind Menon, Samuel Rash, et al. Apache hadoop goes realtime at facebook. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 1071–1080. ACM, 2011.
11. Natural Resources Defense Council. America's data centers consuming and wasting growing amounts of energy.
12. Jeffrey Dean and Sanjay Ghemawat. Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.
13. Yuri Demchenko, Paola Grosso, Cees De Laat, and Peter Membrey. Addressing big data issues in scientific data infrastructure. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, pages 48–55. IEEE, 2013.
14. Marco Dorigo and Mauro Birattari. Ant colony optimization. In *Encyclopedia of machine learning*, pages 36–39. Springer, 2010.
15. Eugen Feller, Louis Rilling, and Christine Morin. Energy-aware ant colony based workload placement in clouds. In *Proceedings of the 2011 IEEE/ACM 12th International Conference on Grid Computing*, pages 26–33. IEEE Computer Society, 2011.
16. M. Ficco and F. Palmieri. Introducing fraudulent energy consumption in cloud infrastructures: A new generation of denial-of-service attacks. *Systems Journal, IEEE*, PP(99):1–11, 2015.
17. Yongqiang Gao, Haibing Guan, Zhengwei Qi, Yang Hou, and Liang Liu. A multi-objective ant colony system algorithm for virtual machine placement in cloud computing. *Journal of Computer and System Sciences*, 79(8):1230–1242, 2013.
18. Bogdan Ghit, Mihai Capota, Tim Hegeman, Jan Hidders, Dick Epema, and Alexandru Iosup. V for vicissitude: The challenge of scaling complex big data workflows. In *Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on*, pages 927–932. IEEE, 2014.
19. Hadi Goudarzi and Massoud Pedram. Energy-efficient virtual machine replication and placement in a cloud computing system. In *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, pages 750–757. IEEE, 2012.
20. Enhanced Intel. Speedstep® technology for the intel® pentium® m processor, 2004.
21. George V Iordache, Marcela S Boboila, Florin Pop, Corina Stratan, and Valentin Cristea. A decentralized strategy for genetic scheduling in heterogeneous environments. In *On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE*, pages 1234–1251. Springer, 2006.
22. Michael Isard, Mihai Budiu, Yuan Yu, Andrew Birrell, and Dennis Fetterly. Dryad: distributed data-parallel programs from sequential building blocks. In *ACM SIGOPS Operating Systems Review*, volume 41, pages 59–72. ACM, 2007.
23. Atefeh Khosravi, Saurabh Kumar Garg, and Rajkumar Buyya. Energy and carbon-efficient placement of virtual machines in distributed cloud data centers. In *Euro-Par 2013 Parallel Processing*, pages 317–328. Springer, 2013.
24. Joanna KołOdziej and Samee Ullah Khan. Multi-level hierarchic genetic-based scheduling of independent jobs in dynamic heterogeneous grid environment. *Information Sciences*, 214:1–19, 2012.
25. Joanna Kołodziej, Samee Ullah Khan, and Fatos Xhafa. Genetic algorithms for energy-aware scheduling in computational grids. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2011 International Conference on*, pages 17–24. IEEE, 2011.
26. Joanna Kolodziej, Magdalena Szmajduch, Tahir Maqsood, Sajjad Ahmad Madani, Nasro Min-Allah, and Samee U Khan. Energy-aware grid scheduling of independent tasks and highly distributed data. In *Frontiers of Information Technology (FIT), 2013 11th International Conference on*, pages 211–216. IEEE, 2013.
27. Joanna Koodziej, Samee U Khan, Magdalena Szmajduch, Lizhe Wang, Dan Chen, et al. Genetic-based solutions for independent batch scheduling in data grids. 2013.
28. Roman Kopetzky, Markus Günther, Natalia Kryvinska, Andreas Mladenow, Christine Strauss, and Christian Stummer. Strategic management of disruptive technologies: a practical framework in the context of voice services and of computing towards the cloud. *International Journal of Grid and Utility Computing*, 4(1):47–59, 2013.
29. Lawrence T. Kou and George Markowsky. Multidimensional bin packing algorithms. *IBM Journal of Research and development*, 21(5):443–448, 1977.
30. Min Yeol Lim, Allan Porterfield, and Robert Fowler. Softpower: fine-grain power estimations using performance counters. In *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, pages 308–311. ACM, 2010.
31. Ching-Chi Lin, Pangfeng Liu, and Jan-Jan Wu. Energy-aware virtual machine dynamic provision and scheduling for cloud computing. In *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, pages 736–737. IEEE, 2011.
32. P Maciel, G Callou, E Tavares, E Sousa, B Silva, et al. Estimating reliability importance and total cost of acquisition for data center power infrastructures. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 421–426. IEEE, 2011.

33. Yousri Mhedheb, Foued Jrad, Jie Tao, Jiaqi Zhao, Joanna Kołodziej, and Achim Streit. Load and thermal-aware vm scheduling on the cloud. In *Algorithms and Architectures for Parallel Processing*, pages 101–114. Springer, 2013.
34. Microsoft. Microsoft azure cloud computing platform, 2015.
35. Andreas Mladenow, Natalia Kryvinska, and Christine Strauss. Towards cloud-centric service environments. *Journal of Service Science Research*, 4(2):213–234, 2012.
36. Christoph Mobius, Waltenegus Dargie, and Alexander Schill. Power consumption estimation models for processors, virtual machines, and servers. *Parallel and Distributed Systems, IEEE Transactions on*, 25(6):1600–1614, 2014.
37. Eugen Molnar, Natalia Kryvinska, and M Greguś. Customer driven big-data analytics for the companies servitization.
38. Catalin Negru, Mariana Mocanu, Costin Chiru, Aurelian Draghia, and Radu Drobot. Cost efficient cloud-based service oriented architecture for water pollution prediction. In *Intelligent Computer Communication and Processing (ICCP), 2015 IEEE International Conference on*, pages 417–423. IEEE, 2015.
39. Catalin Negru, Mariana Mocanu, and Valentin Cristea. Impact of virtual machines heterogeneity on data center power consumption in data-intensive applications. In *Adaptive Resource Management and Scheduling for Cloud Computing*. Springer, 2015.
40. Ewa Niewiadomska-Szynkiewicz, Andrzej Sikora, Piotr Arabas, Mariusz Kamola, Marcin Mincer, and Joanna Kołodziej. Dynamic power management in energy-aware computer networks and data intensive computing systems. *Future Generation Computer Systems*, 37:284–296, 2014.
41. Venkatesh Pallipadi. Enhanced intel speedstep technology and demand-based switching on linux. *Intel Developer Service*, 2009.
42. Rina Panigrahy, Kunal Talwar, Lincoln Uyeda, and Udi Wieder. Heuristics for vector bin packing. *research. microsoft. com*, 2011.
43. Padmanabhan Pillai and Kang G Shin. Real-time dynamic voltage scaling for low-power embedded operating systems. In *ACM SIGOPS Operating Systems Review*, volume 35, pages 89–102. ACM, 2001.
44. Florin Pop, Ciprian Dobre, Valentin Cristea, Nik Bessis, Fatos Xhafa, and Leonard Barolli. Deadline scheduling for aperiodic tasks in inter-cloud environments: a new approach to resource management. *The Journal of Supercomputing*, 71(5):1754–1765, 2015.
45. George Prekas, Mia Primorac, Adam Belay, Christos Kozyrakis, and Edouard Bugnion. Energy proportionality and workload consolidation for latency-critical applications. In *Proceedings of the Sixth ACM Symposium on Cloud Computing*, pages 342–355. ACM, 2015.
46. Andrei Sfrent and Florin Pop. Asymptotic scheduling for many task computing in big data platforms. *Information Sciences*, 2015.
47. Mohsen Sharifi, Hadi Salimi, and Mahsa Najafzadeh. Power-efficient distributed scheduling of virtual machines using workload-aware consolidation techniques. *The Journal of Supercomputing*, 61(1):46–66, 2012.
48. S. Sotiriadis, N. Bessis, A. Anjum, and R. Buyya. An inter-cloud meta-scheduling (icms) simulation framework: Architecture and evaluation. *IEEE Transactions on Services Computing*, PP(99):1–1, 2015.
49. Stelios Sotiriadis, Nik Bessis, and Nick Antonopoulos. Towards inter-cloud schedulers: A survey of meta-scheduling approaches. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2011 International Conference on*, pages 59–66. IEEE, 2011.
50. Apache Spark. Lightning-fast cluster computing, apache spark, 2015.
51. Apache Storm. Storm, distributed and fault-tolerant real-time computation. 2014.
52. Mark W Welker and One AMD Place. Amd processor performance evaluation guide, 2003.
53. Peng Xiao, Zhigang Hu, Dongbo Liu, Guofeng Yan, and Xilong Qu. Virtual machine power measuring technique with bounded error in cloud environments. *Journal of Network and Computer Applications*, 36(2):818–828, 2013.