Energy-Efficient Nature-Inspired Techniques in Cloud Computing Datacenters

Mohammed Joda Usman[1,7*], Abdul Samad Ismail[1], Gaddafi Abdul-Salaam[2], Hassan Chizari[3], Omprakash Kaiwartya[4], Abdulsalam Yau Gital[5], Muhammed Abdullahi[6], Ahmed Aliyu[1, 7] and Salihu Idi Dishing[6]

[1] Department of Computer Science, Universiti Teknology Malaysia, 81310 Skudai Johor, Malaysia (umjoda@gmail.com, absamad@utm.my, ahmedaliyu8513@gmail)

[2] Deparment of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana (gaddafi.ict@knust.edu.gh, gaddafi.ict@gmail.com)

[3] School of Computing and Technology, University of Gloucestershire, Cheltenham, UK. (hchizari@glos.ac.uk)

[4] Department of Computer and Information Sciences, Northumbria University United Kingdom (omokop@gmail.com)

[5] Department of Computer Science, Abubakar Tafawa Balewa Bauchi, 740272 Bauchi, Nigeria (asgital@yahoo.com)

[6] Department of Computer Science, Ahmadu Bello University Zaria, 81007 Kaduna, Nigeria (sidishing@abu.edu.ng)

[7] Department of Mathematics, Bauchi State University Gadau, PMB 068 Bauchi State, Nigeria

***Corresponding Email**: umjoda@gmail.com

Abstract

Cloud computing connotes the systematic delivery of computing resources as services to a wide range of users via the Internet. One form of Cloud computing, Infrastructure as a Service (IaaS), ensures the availability of the resources in the form of Virtual Machines (VMs). Such services are leased to users based on demand and are paid for on a pay-per-use basis. This helps to reduce the cost of running the computing needs of the users. Usually, the VMs are ran on datacenters comprise several computing resources that consume lots of energy, and causing hazardous levels of carbon emissions into the atmosphere. Several researchers have proposed various energy-efficient methods of reducing the energy consumption of the datacenters. Nature has been a cause of inspiration and had a solution to all problem. Therefore, this paper presents a comprehensive review of the state-of-the-art, Nature-Inspired algorithms that have been used in solving the energy issues in the Cloud datacenters. We have categorized all the methods considered into three main techniques; virtualization, consolidation, and energy-awareness. Moreover, we reviewed the different methods in terms their goals, methods, advantages, and limitations. We then compared the nature-inspired algorithms based on their features to indicate their utilization of resources and their levels of energy-efficiency. Finally, we have suggested the potential research directions in this research field. We believe that this review work will be of interest to researchers and professional in Cloud computing datacenters in their quest to providing better energy-efficient methods to address the energy consumption issues of the Cloud datacenters.

.

Keywords: Cloud Computing. Datacenters. Energy-Efficiency. Nature-Inspired Techniques.

1. Introduction

Cloud computing has rapidly developed as an accessible model for providing ICT infrastructure over the last few years. Its broader acceptance and virtualization technologies have contributed to the formation of large-scale datacenters that offer Cloud services. These services are provided in different forms which include Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a service (SaaS) [1, 2]. Cloud computing provides on-demand and elastic computing resources on the pay-per-use basis; thereby, reducing economic cost and increasing the convenience of usage. Figure 1 shows the characteristics of Cloud computing model and their respective components. Energy-efficiency are issues in the different domain of computer science and application, including Internet of things, smart city application and wireless sensor networks [3, 4]. The ever-increasing demand for Cloud services comes with higher energy consumption of the Cloud datacenters. Datacenter(s) are composed of collection(s) of well-structured physical machines including servers and networking equipment on which VMs perform their function. Moreover, in many instances, the datacenter infrastructures are over-provisioned to guarantee absolute service reliability and availability [5]. Furthermore, more than 30% of the physical machines (PM) or servers within Cloud datacenters are usually idle, and often utilizes 10-15% of their resource capacity [6]. Underutilization of the IaaS resources due to inefficient resource management technique is the reason for the high energy consumption and resource under-utilization [7, 8]. The consumed energy is equivalent to 0.01% of global energy consumption which can power 200,000 houses [9]. As a result, the infrastructure of the datacenters produce significant carbon emissions (CO_2), which poses threats to human life and the environment [2].

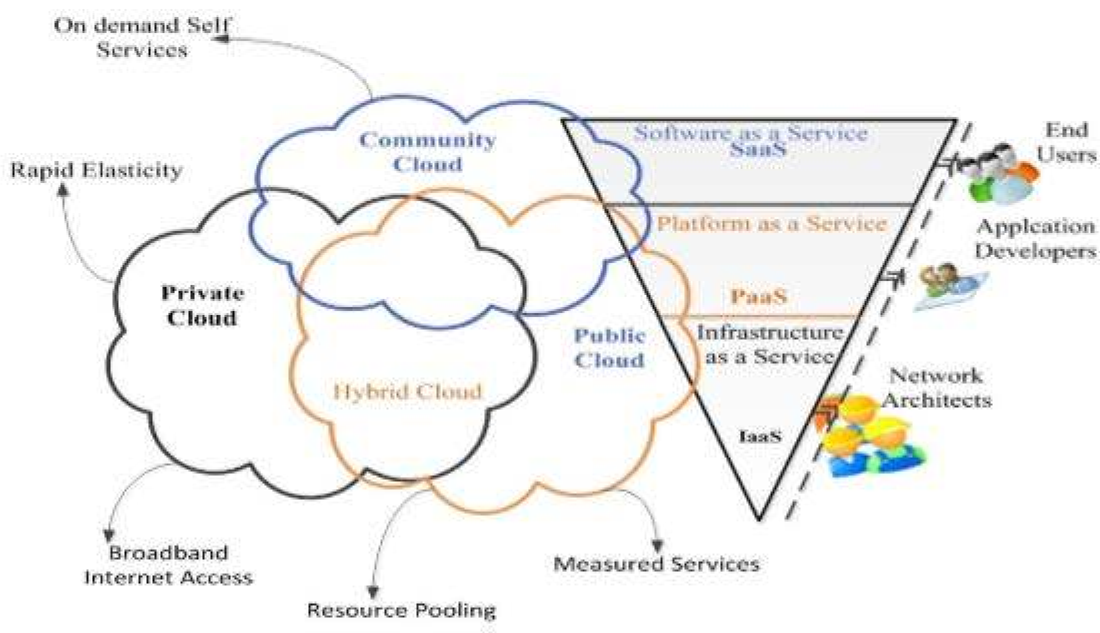

Fig 1. Classification and Models of Cloud Computing

Conventionally, energy efficiency in Cloud datacenters is an NP-hard problem, and various methods have been proposed to address the problem [10]. Amongst them are Nature-Inspired algorithms. These algorithms are usually adjusted to deliver advanced solutions energy efficient issues. The replication of nature produces the algorithms and thus, are believed to be more efficient compared to other methods [11-13]. Therefore, the Nature-Inspired algorithms are required to solve the energy consumption issues of the Cloud datacenters. Several researchers have used them to propose various resource scheduling and allocation technique to improve the energy-efficiency of datacenters.

This paper presents the state-of-the-art methods of Nature-Inspired algorithms that are used in addressing energy consumption issues of Cloud datacenters. It discusses the techniques used and points out their benefits and limitations. Then future research directions in the field are also indicated. This study can enhance the understanding of researchers in their quest to develop and improved energy-efficient techniques. Many studies such as Beloglazov, Buyya [14], Jing, Ali [15], and Kaur and Chana [16] have reviewed energy efficient resource management techniques in Cloud computing datacenters. However, those previous studies considered both Nature-Inspired and non-Nature-Inspired energy-efficient scheduling algorithms in Cloud datacenters. Furthermore, all these research works did not explore in details the Nature-Inspired techniques. To the best of our knowledge, there is currently no comprehensive state-of-the-art review work that focuses on Nature- Inspired energy-efficient techniques as presented in this study. Therefore, this review work differs from the previous studies in that respect as discussed in Section 2.

The rest of this paper is organized as follows: Section 2 presents comparisons of other related reviews and surveys. Section 3 explains the Nature-Inspired algorithms including energy management in Cloud datacenters, energy saving approaches, energy-efficient metrics and Nature-Inspired Optimization approaches. Then, Section 4 discusses the future research direction. Finally, Section 5 concludes the paper.

2. Related Work

This section discusses about the existing survey work in Cloud computing. It presents the similarities and differences between this work and the existing ones. And explains the issues that still need further research.

A review of resource allocation techniques was conducted by Madni, Latiff [17], and Madni, Latiff [18]. The paper focused on the issues of resource scheduling in IaaS Cloud computing environment. It explored resource scheduling schemes and algorithms used by existing works and presented them regarding the problems they solved, schemes and parameters that they used in evaluating their methods. Also, the paper studied the schemes and made a comparative analysis of the parameters used, pointing out their advantages and limitations. In the end, the authors observed that most of the existing schemes failed to incorporate some essential parameters (workload, average resource utilization, and flowtime) in their methods, and urged for improvements of the existing schemes. However, their review was not based on energy-efficient Nature-Inspired optimization approaches.

Kalra and Singh [19] reviewed five metaheuristic techniques: Ant Colony Optimization (ACO) Genetic Algorithm (GA), Particle Swarm Optimization (PSO), League Championship Algorithm (LCA) and BAT algorithm based on scheduling technique. The authors explored these algorithms and provided extensive analysis and comparison of the techniques used in the Cloud computing environment. They presented open challenges and suggested ways of improving the solutions. The solutions quality produced by the algorithms are defined by speed or convergence, initial population generation, transition operator and energy conservation using hybridization concept. Although the work is on Nature-Inspired algorithms, it does not focus on energy management of a datacenter IaaS.

Kaur and Chana [16] presented a survey on Cloud datacenter energy-efficiency. The survey explored different techniques that are proposed to either reduce or overcome the power wastage in Cloud computing. They studied Live VM migration, multi-core architectures, power and thermal-aware, and consolidation technique. According to the authors, virtualization has been an essential technique for efficient resource management and energy utilization of datacenters. The survey emphasized on solutions that are software-based which can be easily integrated into existing infrastructures. Even though the authors have focused on energy-efficient resource management, the emphasis was on heuristics algorithms and not on Nature-Inspired optimization approaches.

Hameed, Khoshkbarforoushha [20] present a survey and taxonomy of energy-efficient resource allocation. The survey focuses on resource management. It covers issues such as energy-aware resource adoption, energy optimization objective functions and their allocation techniques. The authors identify open challenges related to the resource allocation by outlining the problems based on hardware and software techniques. Moreover, they presented the advantages and limitations of the techniques which were studied and also summarized them. However, this research work considered only the energy-efficient resource allocation technique that combines all the classes of algorithms. The authors do not discuss the Nature-Inspired algorithms that are used to improve resource utilization and the energy consumption of datacenters.

A taxonomy of Evolutionary-Inspired Solutions for Energy Management in Green Computing was conducted by Kołodziej, Khan [21]. The paper focused on inspired evolutionary solutions for static and dynamic energy management towards energy-aware green computing. The research work explores resource management method in Cloud datacenters with informative summaries and analysis of the evolutionary approaches. The algorithms are classified into single and multi-population based Genetic optimization techniques that work in a dynamic Cloud environment. The authors acknowledge that the evolutionary algorithms are not popular among other algorithms. But their performance as reported by many authors has made it possible in reducing the energy consumed by the systems that make significant progress in a Green computing. Even though, the authors considered the Evolutionary Algorithm, which is also Nature-Inspired. There is need to study other Nature-Inspired types that are used in realizing energy-efficient resource management in datacenters.

Detailed studies of energy efficiency optimizations at multiple levels in datacenters have been conducted by Beloglazov, Buyya [14]. The authors above present a survey and taxonomy of power management schemes based on hardware and software application. However, virtualization which is the underlying part of cloud computing has not been explored in detail. The authors focus more on heuristics algorithms that use workload consolidation of physical resources to reduce the energy consumption of the datacenters. However, the survey is not on Nature-Inspired based approaches.

3. Energy efficiency in Cloud datacenter: - Nature-Inspired technique perspective

3.1. Nature-Inspired algorithms: - Cloud datacenter usage

Nature-inspired algorithms refers to algorithms that mimic phenomena from physics, chemistry, and or biology. The algorithms are categorized into Evolutionary Intelligence, Swarm Intelligence, and Bio-Intelligence as shown in Figure 2. The term derived its foundation through biological components of nature such as humans, animals, and environment; which could be self-optimizing, self-healing, self-learning, and self-processing [11, 12].

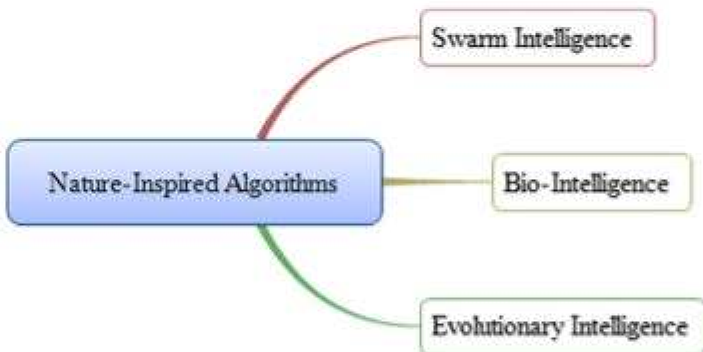

Fig 2. Classification of Nature-Inspired Algorithms

Thus, it is anticipated that computers enabled with their kind of intelligence, can learn to adapt to the changing complexities as nature does in solving complex problems. In fulfilling this goal, algorithms need to adopt the techniques and features from nature to become more efficient and effective [11]. Nature-Inspired algorithms have found applications in many areas. These include optimization problems, resource scheduling /allocation, and load balancing for optimal search solutions. They have been proven to be more advantageous compared to non-nature inspired algorithms. They are considered to be faster in solving complex problems. By applying the Nature-Inspired algorithms in various application settings, new dimensions for scientific research are open for exploitation.

For example, Evolutionary Intelligence which is a Genetic Algorithm (GA) was inspired by Darwin theory of natural selection that is based on survival of the fittest candidate in a given environment [22]. These algorithms begin with a population (set of solutions) which tries to survive in an environment (defined with fitness evaluation). The parent population transmits some of their characteristics of adaptation to the environment to their children through various mechanisms of evolution such as genetic crossover and mutation. The process continues over some generations (iterative process) till the solutions are found to be most suitable for the environment. Swarm intelligence is the group of natural metaheuristics inspired by 'collective intelligence.' The collective intelligence is built up through a population of homogeneous agents interacting with each other and with their environment. An example of such intelligence is found among colonies of ants, flocks of birds, schools of fish, etc. Engelbrecht [23] highlighted the fundamentals and developments in swarm intelligence algorithms for solving numerous real-life optimization problems. The Bio-Intelligence algorithms are mainly derived from living phenomena and behavior of biological organisms. The intelligence associated with Bio-Inspired algorithms are decentralized, distributed, self-organizing and adaptive [12].

Most resource management problems in Cloud computing are faced with large-scale solution space and high computation complexity. Nature-Inspired algorithms can achieve efficiency, by reducing the solution space and leveraging higher-level heuristics, such as genetic algorithm (GA), simulated annealing algorithm (SA), PSO and ACO. In some cases, Nature-Inspired algorithms can find better optimal solution compared with other classes of algorithms. In others, however, it may only return a local optimum in the space searched. Disturbance, such as crossover and mutation in GA, restart in SA, are usually employed to avoid being trapped in the local optimum. In PSO and ant colony optimization, the scheme always follows the current best solution aided by local information of an individual particle to avert local optima trap. The current literature review shows that Nature-Inspired algorithms are used to achieve both single and multi-objective solutions.

3.2 Energy management in Cloud computing datacenter

In energy efficient management, it is necessary to distinguish between Power and Energy. Also, it is important to show their relationships in Cloud datacenter context. That is because decreasing power consumption of a computing device or its component does not necessarily result in overall reduction of energy consumption [24]. Power is the rate at which electricity transfers electric circuit, measured in watts, which is equal to one joule [25]. In essence, power defines how electricity is immediately transferred and or used by a system. On the other hand, energy is the total electricity transferred over a period of time. Additionally, power can also be defined as a function of time, P(t), and energy is the integral of this function Koomey [26].Thus, to reduce energy consumption, it is necessary to reduce the power consumption [24]. However, reducing the power consumed for a limited time does not necessarily reduce the overall energy consumption of the computing components in a datacenter. In this study, the energy management approaches used at a various level of Cloud datacenters are shown in Figure 3.

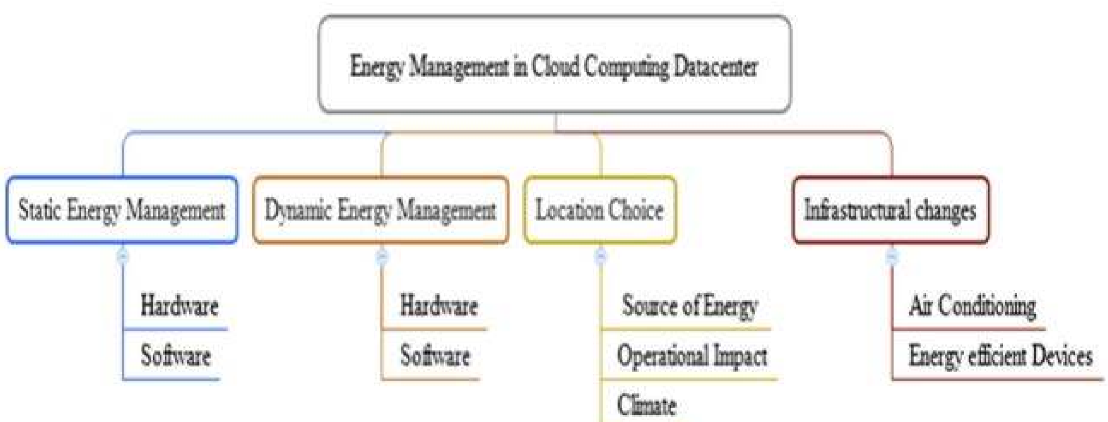

Fig 3. Classification of Energy management in cloud datacenter

Another aspect of energy management for Cloud datacenters is the geographical differences in the energy cost and how clean the source of the energy is. Moreover, geographical location and IaaS changes are other factors worthy of consideration in managing the energy consumption of a datacenter and are briefly summarized in the next paragraph. The energy management techniques are divided into four categories: Static Energy Management (SEM), Dynamic Energy Management (DEM), Location Choices, and Infrastructural Changes [2].

1. Static Energy Management (SEM): From the hardware perspective, SEM includes all optimization methods that are applicable at design time, usually on the circuit, logic and at the system architecture levels. However, even for perfectly designed hardware, poor software design can lead to low performance and loss of power.

2. Dynamic Energy Management (DEM): The SEM techniques are not scalable to address run-time adaptation of datacenters in response to workload changes. But the DEM method is scalable and thus, optimizes energy consumption at the software level. This technique has a mechanism to monitor dynamic characteristics of system's state and adapts according to the current workload requirement. However, DEM requires appropriate adaptation of each component, and understanding of their interaction between individual components when they operate as a system.

3. Location Choice: Geographical location is also considered because the datacenters generate a lot of heat that must be cooled. Likewise the carbon emission of the datacenter will depend on the source of energy of the area where the datacenter is established. Thus, the bigger the datacenter, the more carbon emission because of the number of concurrent users which cannot be suspended or reduced due to energy consumption. Based on these factors we can say that execution of workload in different datacenters located in different geographical location can also vary. Likewise the climate change may also influence the ventilation system that the datacenter may use. For example, if the datacenter is placed in cold area like Finland or Moscow, there is no need to use air conditioning but rather uses the outside breeze in cooling the datacenter infrastructure.

4. Infrastructural Changes: Finally, changing the infrastructure into energy-efficient equipment represent another way of managing the datacenter energy. The datacenter's owners' may decide to change their servers, improved cooling systems, and to migrate to advance software's that are programmed to keep the datacenter architecture at optimal utilization and this in turn, leads to low energy consumption.

Decreasing energy consumption of the Cloud datacenter results in a more sustainable and energy efficient Cloud Computing operations. In the following sub-sections, the practical approaches for reducing the energy consumption of datacenters are discussed.

3.2.1. Active versus Idle Low-power based approach

Under this category, datacenter energy is decreased in two ways, which is active and idle states. A device in active state performs useful work otherwise it will be in a sleeping state [27]. It can also be in the low active state if it is operating at a slow rate with a low-power compared to the active state. On the other hand, in an idle state, computing components do not perform any work. An example of active power mode is dynamically adopting the system processor frequency. During off-peak, the frequency can be changed to constant or discrete steps; a method described as dynamic voltage frequency scaling (DVFS). This technology formerly available only on laptop CPU to align their working frequency and power consumption to conserve battery life is becoming standard on new HPC nodes and servers in Cloud datacenters. Under DVFS, performance states, also called P-states, define the frequencies at which a processor can operate. The Pstates-P0, P1, P2. . . Pn, where n is processor dependent can be examined to prevent energy wastage. For instance, at P3, a CPU will take longer time and use less energy than CPU at P1 [28].

Kessaci, Mezmaz [29] presented energy-aware scheduling that uses evolutionary algorithm combined with DVFS to minimize the datacenter energy consumption. The method used to allocate tasks to processors without violating the precedence constraint and the application approach that is used by the energy-conscious scheduling. Kliazovich, Bouvry [30] explored DVFS technique to schedule datacenter resources. The methodology combines the energy consumption of the network and server components taking into consideration the SLA, traffic demand and the total energy consumption of the datacenter.

Meisner, Gold [31] use this technique to provide a clear solution for choosing one state instead of multiple states at a time with different performance, transition time, and energy consumption. The DVFS reduces the power consumption by tens of watts and on a single server component which is the CPU [32, 33]. DVFS technique is not only employed to save energy for a single server and its components but also for the network resource and other related communication components in the datacenter to attain green computing [34]. Similarly, Jiang, Xu [35] utilized the Redundant Links Algorithms (RLA) to detect the highest links that are in sleeping links and reverse those links to realized energy saving for the network resources in the datacenter. Kim, Beloglazov [36] also used power-aware DVFS scaling for energy management in Cloud datacenters. Some tasks scheduling algorithms proposed by Sharma and Reddy [37] and Yassa, Chelouah [38] used DVFS technique which allows for energy savings when the PM is not fully utilized. Despite the advancements in this technique, the energy consumption of Cloud datacenters has not dropped because the technique (DVFS) is limited to only the CPU. Therefore, the focus has moved from the hardware component to new techniques that are currently being implemented by datacenter administrators to optimize both hardware and software components of IaaS.

3.2.2. Energy - Aware hardware potentiality based approach

In this category, the energy reduction focuses on individual components of a system such as the CPU, memory, disk and network component of the Cloud datacenter resources. Their efficiencies can be improved by manufacturers that provide hardware optimizations. Many research studies such as Gabrel Torres [39], Snowdon, Le Sueur [40], Ousterhout, Agrawal [41], Koomey [42], Hähnel, Döbel [43] Eom, Choi [44] and Jiang, Xu [45] uses this approach to improve their energy efficiency and performance.

3.3.3. Advancement in software development based approach

Nowadays, software applications, drivers, and module kernel are developed with energy consumption consciousness. They provide management support that allows users access to the operating state of the device and their energy consumption [46]. Michael and Krieger [47] exploit different versions of windows and Linux by running the same applications on them and discovered that the power consumption under the distinct versions shows non-negligible variations. Operating System (OS) is the central software that drives the system functionality on any computing devices. OSs have different power consumptions that can be optimized to consume less energy based on their type and version [31].

3.2.4. Consolidating IT resources based approach

Another approach for saving energy is consolidation and sharing of a particular resource such as power supply, memory, disk drive or CPU instead of using a number of them per server racks. This method reduces the energy consumption in the datacenter when fewer machines are utilized [48]. In this regard, datacenter networks have received considerable attention [49] and are therefore consolidated, tasks of underutilized devices can be rerouted to other devices to make them operate at full capacity using minimum criticality technique [4]. Hence, only few equipment is used, resulting in less power consumption. Similarly, replacing old servers with blade servers which consume about 10% less power than conventional servers are also help to save energy [50, 51]. In this regard, efficient consolidation does not only enable carrying the maximum workload on least number of resources but also keeps each resource (CPU, disk, network, etc.) at maximum utilization.

3.3. Metrics for Energy Efficiency in Cloud Datacenter

Energy efficiency metrics are the scale of measurement used to assess the operational condition of Cloud datacenters. The metrics determine energy efficiency for each level of Cloud architecture (infrastructure, virtualization, and application), and the interrelation among them. Power consumption and energy-efficiency are fundamental factors in the initial design and day-to-day management of computer systems [52]. Energy efficiency has become a significant metric that is increasingly implemented to evaluate and measure the energy utilization of devices installed in datacenters [53]. The Green Grid realizes the importance of establishing metrics for datacenter sustainability. Ideally, these metrics and their related methods help organizations to determine if the datacenter is optimal before demanding a new datacenter [53, 54]. Therefore, to manage the resources of the datacenter efficiently, there is need to measure the efficiency of a datacenter regarding energy consumption which has affected the designing and development of resource optimization metrics by the Green Grid Association (GGA). The GGA introduces the Power Usage Effectiveness (PUE), Data center Infrastructure Efficiency (DCiE), Carbon Usage Effectiveness (CUE) and Data Center Energy Productivity (DCeP) to serve as yardsticks for evaluation as well as improvement of datacenters performance.

PUE is the most widely datacenter metric use today. The PUE is defined as the ratio of the total power consumption of a datacenter facility (electrical systems, air conditioning, lightning and other related equipment in the infrastructure) to the total power consumption of IT equipment (servers, surge frames, storage, network components, etc.), as expressed by Eq 1.

$$\mathrm{PUE} = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}} \tag{1}$$

Low PUE presumably means higher efficiency because the significant part of the power has been used by IT equipment.

DCiE is the inverse of PUE which is a general metric for measuring the datacenter energy efficiency. This metric is calculated by dividing IT equipment power consumption by facility power, and it is expressed in percentage. Some datacenter operators use these metrics more due to its helpfulness in benchmarking and understanding the power overhead incurred by datacenters as shown by Eq 2.

$$\mathrm{DCiE} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}} \tag{2}$$

To get an advantageous single metric for DCiE the total energy use should be measured (e.g., in kWh) for a period that is longer than the cyclic variation in efficiency, for many facilities, this may be a full year [55].

CUE is another metric used by datacenter operators. It provides detail of certain environment efficiency relative to carbon emission. It is the total carbon emission caused by the cumulative datacenter energy divided by IT equipment energy. It has the same denominator with PUE but the numerator focuses on carbon emission, and it depends on the source of energy used by the datacenters as expressed by Eq 3.

$$\mathrm{CUE} = \frac{\text{Total CO2 Emissions caused by Total Data Center Energy}}{\text{IT Equipment Energy}} \tag{3}$$

The 'Complete CO_2 Emissions' are weighed in kilograms of carbon dioxide ($kgCO_2eq$) per kilowatt-hour (kWh) and 'Total datacenter Energy' is the amount of power used as measured by the utility meter. If your datacenter is running entirely on power-grid electricity, the region-wise government data will give you the numbers [56].

DCeP is also a metric that is being used to evaluate the useful work performed by the datacenter based on the quantity of energy utilization over a period of time. DCeP has been considered as the most effective and efficient method for measuring the whole of datacenter efficiency. This metric can be mathematically shown as in Eq 4.

$$\mathrm{DCeP} = \frac{\text{Total Useful Work was done}}{\text{Total Data Center Energy Consumed Over Time}} \tag{4}$$

The DCeP essentially defines the datacenter as a black box – power goes into the box, heat comes out, data goes into and out of the black box, and a net amount of useful work is done by the black box [57].

All the afore-mentioned metrics were introduced by the Green Grid with the aim of maximizing the energy efficiency of datacenters. However, other metrics that are not mentioned here can be found in the energy efficiency of the distributed system by Zomaya and Lee [58], best practices for energy efficient datacenter by VanGeet, Lintner [59] and energy-efficiency metrics for datacenter by Newcombe [60]. In Table I, we categorize the existing efficiency metrics based on their application areas and their formulae of computations.

Table 1: Summary of energy efficiency metrics and formulae of computation

Name of the Metrics	Computational Formula
Power usage Effectiveness	$\text{PUE} = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}}$
Carbon Usage Effectiveness	$\text{CUE} = \frac{\text{Total CO2 Emissions caused by Total Data Center Energy}}{\text{IT Equipment Energy}}$
Water Usage Effectiveness	$\text{WUE} = \frac{\text{Annual Water Usage}}{\text{IT Equipment Energy}}$
Energy Reuse Factor	$\text{ERF} = \frac{\text{Reuse energy outside of the data center}}{\text{Total Data Center Source Energy}}$
Energy Reuse Effectiveness	$\text{ERE} = \frac{\text{Total Energy} - \text{Reuse Energy}}{\text{Total IT Equipment Energy}}$
Data center Infrastructure Efficiency	$\text{DCiE} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}}$
Data Center Productivity	$\text{DCP} = \frac{\text{Useful Work}}{\text{Total Facility Power}}$
Compute Power Efficiency	$\text{CPE} = \frac{\text{IT Equipment Utilization Energy}}{\text{PUE}}$
Green Energy Coefficient	$\text{GEC} = \frac{\text{Green Energy Consumed}}{\text{Total Energy Consumed}}$
Space, Wattage, and Performance	$\text{SWaP} = \frac{\text{Performance}}{\text{Space} * \text{Power}}$
Data Center Energy Productivity	$\text{DCeP} = \frac{\text{Total Useful Work was done}}{\text{Total Data Center Energy Consumed Over Time}}$

3.4. Nature-Inspired energy-efficient techniques for Cloud datacenter

Several energy-efficient scheduling techniques have been developed using Nature-Inspired algorithms to prevent resource underutilization, which is one of the attributes responsible for incurring high energy consumption [61]. The energy-efficient techniques are classified into two main classes; Non-nature inspired (Heuristics) and Nature-Inspired (Meta-Heuristics). In this work, we focus only on the Nature-Inspired category. Nature-Inspired optimization may be classified as single or multi-objective (SOO or MOO) depending on the nature of the objective function. In the following subsections, we have explained them together with their corresponding techniques.

3.4.1. Optimization approaches for Cloud datacenter

Multi-Objective Optimization (MOO) is becoming more popular in the quest for a solution to real word problems. Nowadays due to its capability to model different scenarios, it has often been used by researchers. MOO has two approaches; to combine individual objective functions into a single composite function, or move all except one objective to the constraint set known as Pareto optimal. The general form of the equation for MOO is expressed in Eq 5.

$$\text{Optimize (Min/Max) } F(x) = \{f_1(x), f_2(x), \dots \dots \dots, f_n(x)\} \quad (5)$$

Subject to $\mathrm{U}\,(X) = 0, \mathrm{Y}\,(X) \geq 0$

The optimized functions (minimize or maximize) are the set of functions $F(x)$, and the vector x is considered to be the independent set of variables. Functions $U(X)$ and $Y(X)$ are the constraints of the model. The solutions solve the objective function even when they are conflicting. That is when minimizing one function may worsen others.

The Single Objective Optimization (SOO) approach finds the "best" solution that corresponds to the minimum or maximum value of a single objective function which groups together all the different objectives into one. This approach is suitable for use if the required parameter(s) of the search is not laborious. Examples of the works that have used SOO include Babukarthik, Raju [62], Quang-Hung, Nien [63] for energy-aware technique; Wu, Tang [64], Wu, Tang [65] for consolidation technique; and Tao *et al.* (2015), Luo, Cao [66] for Virtualization technique. The search for an optimum solution is the only goal in SOO even if there are different optimal solutions. If new results yield better objective function value than previous, the new results are accepted. Mathematically, SOO can be expressed as in Eq 6.

$$Optimize\ (Min/Max)\ f(x) \quad (6)$$

Subject to $\mathrm{U}\,(X) = 0, \mathrm{Y}\,(X) \geq 0$. In Eq 6, the objective function to be optimized (minimized or maximized) is $f(X)$, and vector X is the set of independent variables. The functions $U(X)$ and $Y(X)$ are constraints of the model. The methods that have adopted the aforementioned approaches are discussed in subsequent sections.

3.4.2. Energy-efficient techniques for Cloud datacenters

Energy-efficient resource management techniques are categorized into three according to their deployment mode. These are Virtualization, Consolidation, and Energy-aware. Each technique employs either the MOO or SOO approach to improve resource management and to reduce the energy consumption of the datacenters. We present in the next subsection various works that are based on SOO and MOO formulation.

3.4.2.1. Energy-Aware oriented technique

To reduce energy demand and improve resource utilization, energy-aware techniques have been used to schedule resources in Cloud datacenters as in Mezmaz, Melab [67], Malakooti, Sheikh [68] and Raju, Amudhavel [69]. This technique uses resource scheduling to improve resource utilization resulting in sustainable datacenters. The technique uses GA and Bat Intelligence (BI) to schedule resources for execution in a datacenter, taking into consideration the energy consumption, tardiness, and makespan. Since the scheduling depends on the size and the number of available resource in the datacenter. This method has been proved to be effective. However, it still consumes high energy when the size of the request is large.

Yassa, Chelouah [38] proposed an energy-aware multi-objective approach for workflow scheduling for Cloud datacenter. In the approach, a Multi-objective Discrete Particle Swarm Optimization (MODPSO) combines with DVFS to form a hybrid iterative method that reduces energy consumption. The DVFS decreases supply voltage which lowers the clock frequency of the CPU; thereby reducing the power consumption. Moreover, when the CPU is idle, it goes into sleep mode reducing the supply voltage and clock frequency. This method makes it possible for the CPU to operate at different voltage levels by considering several objectives such as energy consumption, makespan, and cost. The algorithm starts by initializing the positions and velocities of particles. To obtain the position of a particle, the (voltage and frequency) VSL of each resource is randomly initialized, then a Heterogeneous Earliest-Finish-Time (HEFT) algorithm is applied to generate a feasible and efficient solution for minimizing the makespan. The process repeats several times to initialize the positions of all particles of the swarm. Initially, the velocity, *V,* and the best position for each particle, *pBest,* are attributed to the particle itself. The algorithm maintains an External Archive (EA) to store non-dominated particles found after the evaluation process. Subsequently, the new *V* and *pBest* of each particle are calculated after selecting the best overall position in the external archive. Eventually, a mutation is performed, then the particle is evaluated, and its corresponding *pBest* is updated. Once the termination condition is reached, the EA containing a Pareto front is returned as a result. The hybridization of these methods provides a set of pareto solutions called non-dominated solutions, that eventually improve resource utilization and reduces energy consumption. However, this approach focuses only on reducing energy consumption at the internal level of the Cloud datacenter infrastructures; it neglects the logical components, which also consume a considerable amount of energy. Therefore, the approach is partially energy-efficient.

Guzek, Pecero [70] developed an Energy-Aware Multi-objective Algorithms (EAMA) for distributed computing systems to reduce energy consumption. In EAMA, three scheduling algorithms are suggested for solving the issue of heterogeneous multi-processor multi-objective scheduling based on state-of-the-art multi-objective algorithm schemes. A Cellular GA (CGA) is used to optimize the workflow schedule in the presence of clear communications costs on homogeneous infrastructures. It minimizes execution time and energy consumption. A solution is served as a vector of request assignments, which is further processed by a list heuristics. The proposed grouping and crossover intrinsically minimizes communications volume thus indirectly, makespan and energy consumption. A uniform mutation operator drives diversification of the search. The CGA is tested against a standard GA in simulations with real application structures and different communication costs. The CGA consistently outperformed the GA, especially for the communication-intensive workflows. Although the scheduling depends on the task and the number of processors, it minimizes the energy consumed by the datacenter. However, EAMA prevents scheduling of task with independent resources. The proposed technique also appears to be in-efficient for the resources and consequently not suitable for Cloud datacenter application.

Feller, Rilling [70] have proposed Energy-Aware ACO (EA-ACO) based on workload placement to deal with the energy resource management problem. Each ant receives all items (i.e., VMs), opens a bin (i.e., physical machine) and starts assigning the items to the bin. The method uses a probabilistic decision that describes the attractiveness for an ant to select a particular item as the next one to pack in its current bin. Thus, the higher the pheromone and heuristic information associated with a particular item-bin pair, the higher the probability that an ant will choose it. This stochastic nature of the algorithm allows the ants to explore a large number of potential solutions and thus, calculates better placements than the evaluated state-of-the-art Greedy Algorithm. After all, ants have constructed their solutions including the amount of pheromone associated with each item-bin pair is updated to simulate pheromone evaporation and reinforce item-bin pairs which belonged to the better solutions. Unlike previous approaches that consider only a single resource, EA-ACO focuses on multiple resources such as CPU cycles, CPU cores, disk size, RAM size and network bandwidth. The approach has been applied in managing datacenters energy management. However, the technique takes longer time in search of an optimal solution. In this way, it can cause extra energy consumption, thereby, lowering the energy-efficiency.

Liu, Zhan [71] suggest an Energy-Aware VM (EAVM-ACO) placement scheduling method based on ACO. The EAVM-ACO is presented as an alternate way to improving resource utilization and energy efficiency. Figure 3 shows how the method constructs a solution, whereby, the VMs are placed on physical servers. It uses live VM migration from one operational node to the other to find the best node among the overloaded VMs to be migrated.

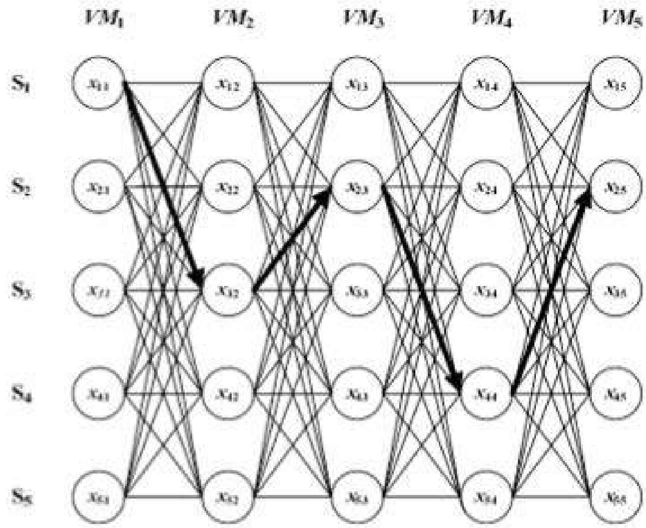

Fig 4. Construction of Solution [72]

Each element x_{ij} represents whether the VM_j is deployed on the server S_i. If x_{ij} is selected, then it is set to 1. Otherwise, it is 0. The total number of steps for an ant to construct a solution is the same as the number of VMs. For each step *j*, the ant selects a physical server to run the current VM_j. There is only one server selected in each column. The VMs selected on row i are deployed on server S_i (*i*=1, 2, 3,… *Mt*). The ant's path $\left(x_{11}, x_{32}, x_{23,}\, x_{44}, x_{25}\right)$ represents a solution *S*= $\{S_{1,}\, S_2, S_3, S_4,\}$, where VM_1 is deployed on server S_1, VM_3 and VM_5 are deployed on S_2, VM_2 is deployed on S_3, and VM_4 is deployed on S_4. S_5 is not used. In EAVM-ACO, the resource utilization is increased plus the energy consumption is reduced. However, there is no performance guarantee while the size of the physical machine increase with the VMs which leads to higher computational time and thus, higher resource wastage with energy consumption.

Energy Aware VM Placement Scheduling in Cloud Computing using Firefly Optimization (FOA) approach is presented by [73]. It makes energy-aware decisions based on the past resource utilization and energy consumption data. This technique tries to migrate the most loaded VM from an active node which satisfies a minimum criterion for energy consumption to another active node that consumes the least energy. It consists of four main parts, A) Selection of a source node, B) Selection of VMs, C) Selection of the destination node, and D) Distance updated based on the flashing behavior of fireflies. The method handles the growing demand for energy and the heterogeneity nature of Cloud datacenters. However, there is no performance guarantee as to the size of the physical machine increase with the VMs which leads to higher computational time. More so, the algorithm has been reported to have weak exploration and exploitation capability due to the excessive use of the turning parameters.

Duan, Chen [74] proposed a PreAntPolicy, an energy-aware scheduling for VMs in heterogeneous Cloud computing systems. The PreAntPolicy is composed of a prediction model that is based on fractal mathematics, and a scheduler from an improved ACO as seen in Figure 4.

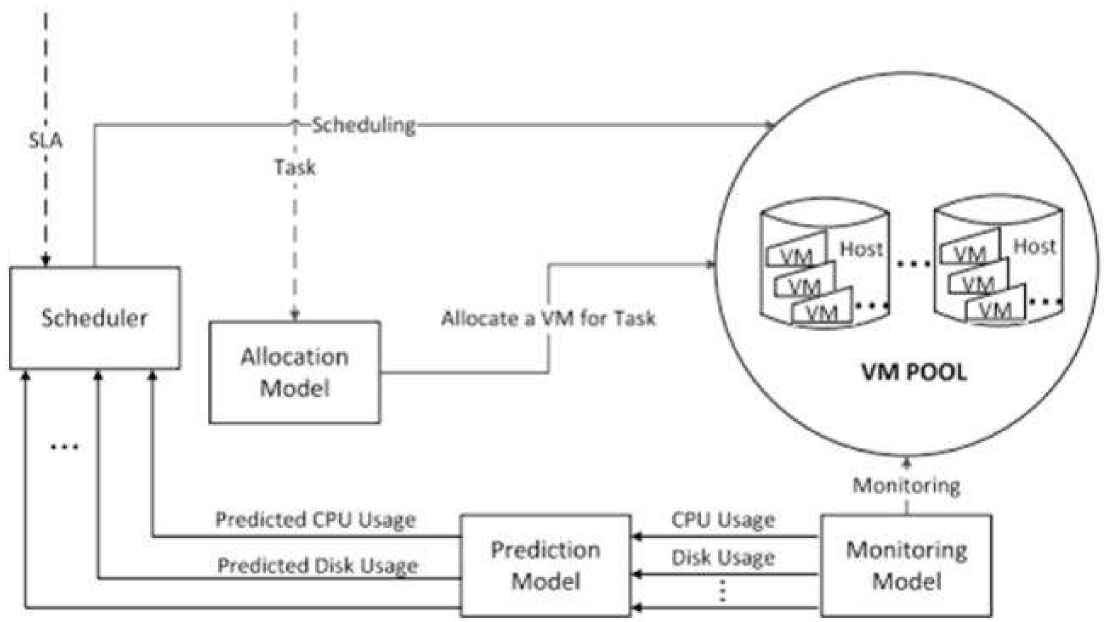

Fig 5. Proposed scheduling approach [74]

The approach extends an excellent dynamic model of resource provisioning applications within a heterogeneous Cloud datacenter that can minimize the overall resource wastage and energy consumption. The prediction model predicts the future resource utilization of a data center through the data collected by the monitoring model. Then the scheduler schedules VMs based on the instructions from the prediction model. The central job of the allocation model is to assign an appropriate VM to a new Cloud request supplied by clients. Although the scheduler supports different types of resource scheduling in the real scenario, each request occupies a single VM. This approach leads to under-utilization of resources, thereby, resulting in high energy consumption.

3.4.2.2. Virtualization oriented technique

Virtualization is another essential and useful component of Cloud computing. The rewards associated with deploying this technology in a Cloud datacenter and its environment are enormous. They include portability of high-level functions, resource sharing, and aggregation of actual physical resources [75]. Evolutionary Intelligence has been the first MOO technique to be used for improving datacenters energy efficiency. The problems have been modeled as NP-hard using GA as the ideal candidate for solutions as in Xu, Zeng [76], Song, Fan [77], Shigeta, Yamashima [78], Gao, Guan [79], Wang, Wang [80], Wang, Wang [81], Ramezani, Lu [82], Yao, Ding [83] and Joda [84].

Phan, Suzuki [85] presented a Green Monster framework and designed Evolutionary Multi-objective Optimization Algorithm (GM-EMOA) to move services dynamically across a federation of geographically dispersed datacenters. Figure 5 shows the interaction between Internet datacenters (IDC) with the Green Monster (GM).

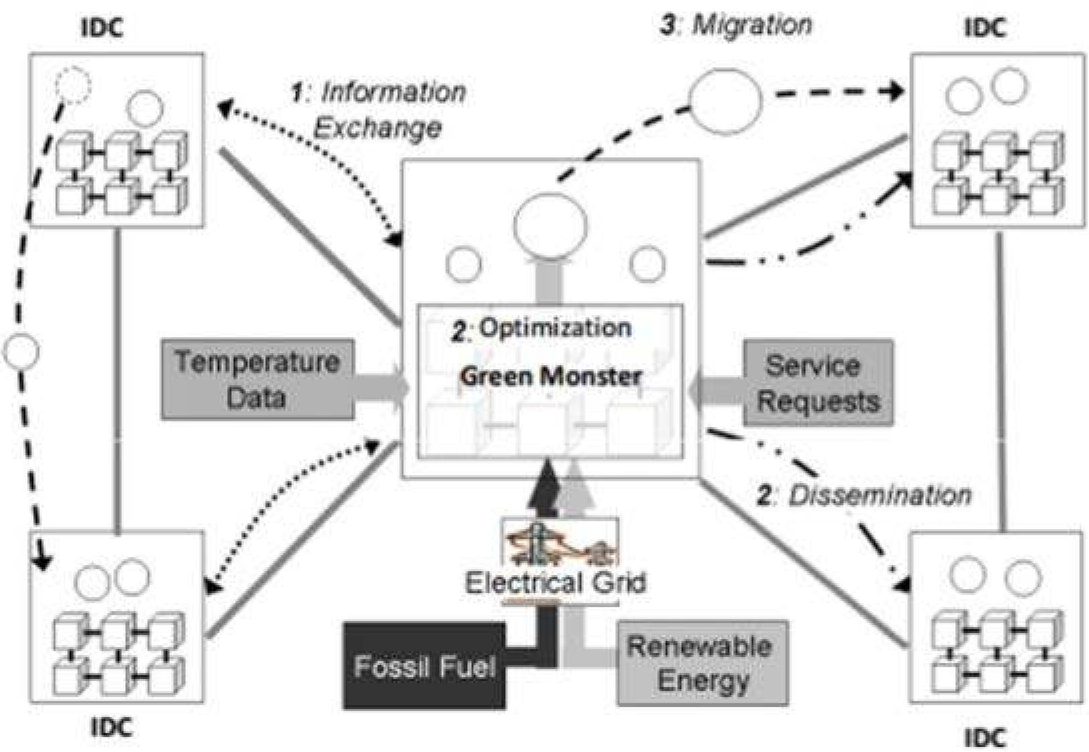

Fig 6. Interaction of IDC and GM [85]

The GM-EMOA algorithm reduces carbon emission without affecting the performance of the system in a green way. GM-EMOA searches the Pareto-optimal solutions by balancing trade-offs between the self-contradicting optimization objectives such as datacenter cooling, renewable energy consumption, and response time. The method implies the response time to users are minimized together with the datacenter cooling that assumes free cooling system when the outdoor temperature is less than the indoor. Also, the renewable energy consumption is maximized. EMOA allows IDC operators to make well-informed optimum decisions for service migration and placement by providing them with a diverse set of approximated Pareto optimal solutions. After a simulation run, one of the non-dominated individuals is chosen as a simulated decision of an IDC operator to perform dynamic service migration and placement. GM-EMOA is environmentally conscious as it is mindful of the CO_2 emissions. However, the slowness in convergence may lead to higher energy-consumption and thus, minimize the energy efficiency.

Shu, Wang [86] present a resource scheduling inspired by Clonal Selection Algorithm (ICSA) to reduce the energy consumption of datacenter IaaS. ICSA has a significant influence on local optima avoidance within a range of a given feasible solution and uses less running time. Once a new request for resource arrives, the system runs the ICSA to balance the resource allocation. The ICSA first changes the mapping relationships between resources and tasks into a binary code as a set of initial population X (0), before discovering the best individual solution X_I^G expressed as in Eq 7.

$$X_i^G = \left(X_{i1,}^G X_{i2,\ldots\ldots,}^G X_{iP}^G\right) \quad (7)$$

Where G is the current generation, i = 1, 2…, s, and s is the population size. Each (antibody) means that a candidate solution is represented by a binary string of bits. The ICSA is applied in resource allocation to deal with the optimization problem, and the affinity function is designed by energy efficiency and makespan. The affinity function can be represented as in Eq 8.

$$aff(x) = e^{min} E_i + min\, MS \quad (8)$$

The ICSA proofs to be more useful in optimizing energy consumption and resource allocation compared with existing algorithms. However, it uses only a few physical machines in its implementation and thus, may not support large datacenters.

Pascual, Lorido-Botrán [87] propose an enhance placement policy with network-aware optimization to assign applications onto servers in cloud datacenter infrastructure. Four different models (i.e., application, workload, datacenter structural and power consumption models) were developed using a web application. These models were implemented on three-layer architecture (Fronted, Business (L1), and Persistence (L2)). The layers define only two classes of web applications to solve the issue. The application model considers three types of request: (1) *p*: it is processed in the business layer, and requires no access to the database; (2) *r*: it requires a query (read operation) to the database; (3) *w*: it requires a write operation on the database. Each request that passes through the application layers has different inter-VM messages and processing times. The structure of the datacenter focuses on interconnection networks built of the Fat-trees topology. Finally, a general model of power utilization of the datacenter resources such as server components and network components is formulated as in Eq 9.

$$E = \{ \quad E_{active} \quad + \frac{\substack{E_{idle} \\ E_{rem.(U_{total}-1)}}}{U_{total}-1} \quad U_{active} \geq 0 \quad (9)$$

The algorithm implements mechanisms that converge rapidly to greater quality placements. The solution is searched using evolutionary techniques based on the selection criteria used to select the best solution for each generation. The method optimizes each of the Pareto set. The authors' demonstrate that allocating applications using the proposed policy results in a high utilization of resources (servers, networking elements) and improves application performance. Although the policy claims to reduce the overall energy consumption, however, there is extra communication overhead on the network and servers due to the dynamic VM migration. This method will lead to resource under-utilization and high energy consumption, thus minimizing the energy-efficiency.

An Energy-efficient scheduling on a green datacenter using the multi-objective co-evolutionary algorithm (OL-PICEA-g) is explored by Lei, Wang [88]. The authors address the energy-efficient scheduling issue for a green datacenter partly maintained by renewable energy and convention energy source. Figure 6 shows how the datacenter schedule workload (computational tasks) and selects the appropriate computing nodes with voltages and clock frequencies to cope with the renewable energy source.

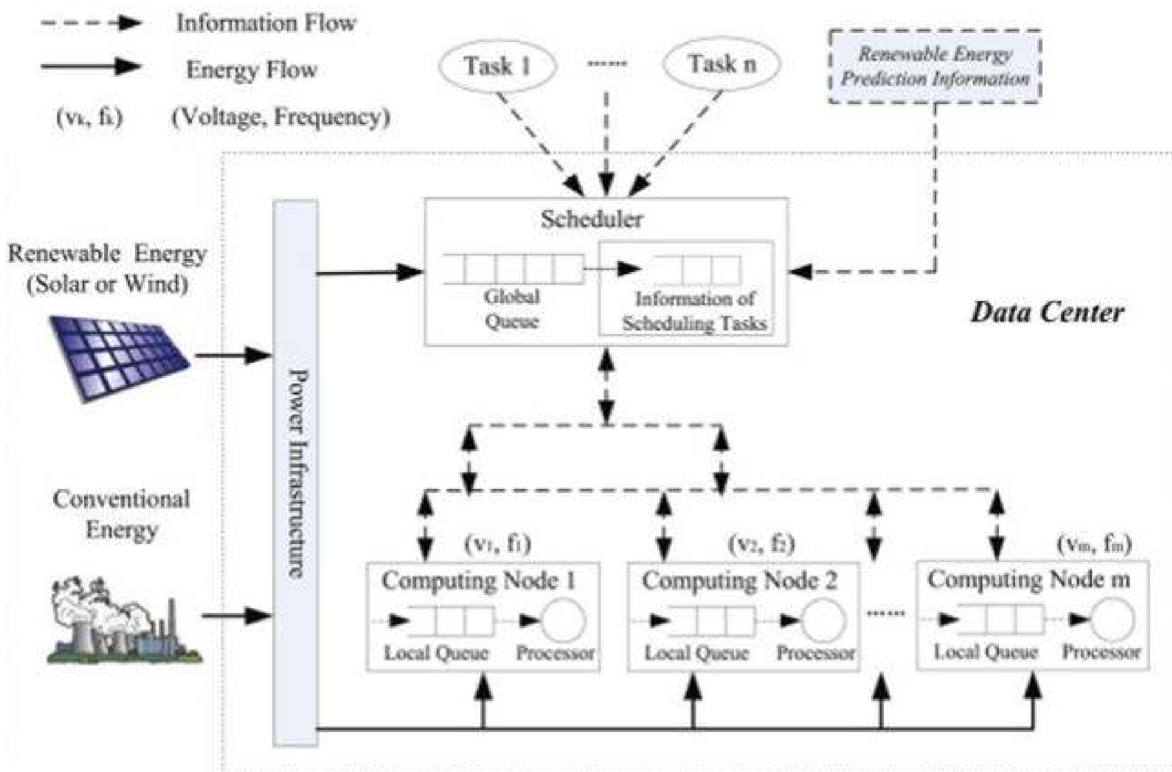

Fig 7. System Architecture [88]

The algorithm assumes a datacenter with *m* homogeneous computing nodes that use two scheduling strategies Green-oriented scheduling strategy (GOSS) and Time-oriented scheduling strategy (TOSS). For the task scheduling, the scheduler obtains prediction information of the renewable energy, collects status information of the datacenter system and uses the scheduling algorithm to dispatch the tasks to suitable computing nodes and decides when to run the tasks. Based on this, a task can be rejected when its deadline cannot be met. By combining these strategies, OL-PICEA-g simultaneously minimizes the makespan of tasks and the total energy consumption. However, only a single task is allowed to execute at a time. The approach avoids the complexity of fine-grained scheduling of multiple resources and leads to resource under-utilization. Consequently, it results in high energy consumption in the datacenters. Thus, even though the OL-PICEA-g encourages the use of renewable energy, it may not be adequately efficient in energy consumption reduction due to the limitation highlighted.

A GA-based Hybrid Optimization (GAHO) model for green cloud computing to establish energy-efficient datacenters is proposed by Rocha and Cardozo [89]. In the model, a VM placement and transport network combines Integer Linear Programming (ILP), GA and network simulations as a strategy to achieve its goal. Three nodes: a source node, sink node, and core are considered. These nodes represent all the traffic with origins and destinations outside the cloud infrastructure. The network is represented as a directed graph where the vertices are network nodes (routers), and edges are communication links connecting the nodes. This algorithm produces non-dominated solutions that allow the cloud provider to choose suitable trade-off regarding energy consumption and network QoS. However, it takes longer time in the presented case study to conduct 100 iterations of the GA. It means that GAHO-ILP may be suitable for organizations that have small hosting servers but not adequate for large datacenters, in terms of energy-efficiency.

Javanmardi, Shojafar [90] and Shojafar, Javanmardi [91] presented a scheduling strategy that uses hybridization concept of fuzzy theory with GA to improve the resource scheduling of the Cloud datacenter. The authors have modified the GA to reduce the time taken to search for optimal solutions with the help of the fuzzy theory. The approach considered execution cost with total execution time of the datacenter resources that result in efficient performance. However, energy consumption has not been considered as it is one of the main parameters to be considered by the datacenter service provider at the IaaS level.

Sharma and Reddy (2015) have combined DVFS and GA to reduce the energy consumption of a datacenter, increased resource utilization and convergence of the solutions. The method map VMs to PMs randomly like mapi = (VMi map to PMj), and coded as a chromosome and used the fitness evolution to fit in VMs on PMs with a small number of the application as shown in Eq 10.

$$fit(i) = \sum_{j=1}^{m} \frac{\sum_{i=1}^{n} pes_i * mips_i * a_{ij}}{PES_i * MIPS_i} \tag{10}$$

$$\text{For } a_{ij} = \begin{cases} a_{ij} = 0 \; if\ VM\ i\ not\ assigned\ to\ PM\ j \\ a_{ij} = 1 \; if\ VM\ i\ assigned\ to\ PM\ j \end{cases}$$

Where $mips_i$ represents millions of instructions per second required by VM i, $MIPS_i$ represents millions of instructions per second executed by server j, pes_i is the number of processing element required by VM i, $MIPS_i$ is the number of processing element required by server i. A new selection operator is applied to generate individual chromosome for the generation and a single crossover is considered. The mutation operation is done on the chromosomes by exchanging the selected PMs and VMs. The process is terminated when the number of iteration is greater than the set value and the result did not reach the set value. These hybrid techniques have the advantage of providing more optimal results than their single ones. However, they increase computational workload due to integrated repairs and local optimization procedure in the hybridization process. They also take demand a longer time to reach the non-dominated solutions and, therefore, are considered to be slow. Although the hybrid algorithm archives its goals, the energy reduction is limited due to the slow convergence.

Moganarangan, Babukarthik [92] propose a Hybrid Algorithm (HA) for reducing energy consumption and makespan in cloud datacenters. The algorithm combines ACO and Cuckoo Search Algorithms (CSA) to reduce the energy consumption of datacenter. The jobs are processed based on arrival from 1 to *n* jobs. After applying transition rules, jobs that arrive are treated. To process a job, it is allocated to VM based upon the arrival. Thus *n* jobs are assigned to the VM_m. Then CSA performs random walk by applying Levy's flight based on the best nest for next generation. Ant moves are performed from *m* to *n*. Update on pheromone trails is applied once a search is performed and the Global update of pheromone is carried out. Then it will list all the necessary resources for the VMs. The performance of HA in search of an optimal solution is faster and energy efficient. Thus, HA performs better, and results in energy consumption reduction of the Cloud datacenter. However, HA focuses only on the energy consumption of the CPU neglecting other components that also consume the energy of the datacenter.

3.4.2.3 Consolidation oriented technique

Consolidation technique employs live VM migration to consolidate VMs systematically so that over utilized PMs can be lessened and migrated to under-utilize PM. The most significant features here is to decide which VM to migrate from over and or under-utilized PMs that will influence the resource utilization and energy-efficiency. The advantage of employing this technique is that it increases efficiency in resource utilization and energy-efficiency of servers in the datacenter.

A multi-objective machine re-allocation algorithm for Cloud datacenters called *GeNePi* is proposed by Saber, Ventresque [93] to maximize resource efficiency. *GeNePi* combines three algorithms to form a hybrid technique in the optimization procedure. Figure 7 shows the selection process of reassigning the resources based on a linear collection of three utility/objective functions.

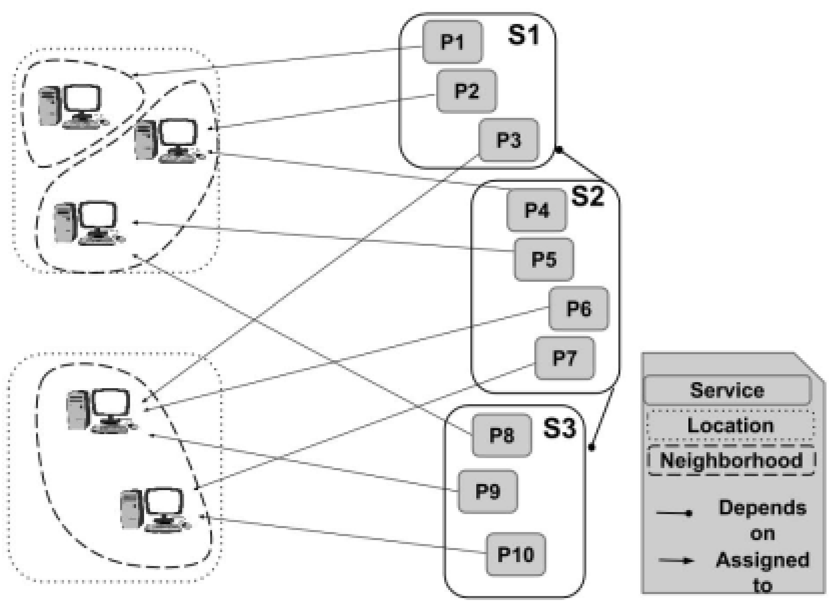

Fig 8. Simple scenario assignment of processes to machines [93]

The solutions are generated by trying to reassign the processes one after another based on a greedy heuristics, which is slightly relaxed to include a random factor. After ranking the processes according to their dependencies and the needs of their resources, they are selected one by one. *GeNePi* makes the reassignment decision of the processes from their initial hosts. The choice of the reassignment of every process is based on a linear combination of the three utility/objective functions (i) reliability of the assignment, (ii) migration cost of the assignment, and (iii) energy consumption. For each process, a set of assignable machines that respect the constraints is computed, and a value of interest is given to each machine by a weighted sum as shown in Eq 11.

$(U_i): f(x) = a_0 + \sum_{i=1}^{3} \curlywedge U_{i,,}$ (11)

where (U_i) is resource, $f(x)$ is the objective function, a_0 is the reassignment number. The equation creates a set of machine with a utility lower than or equal to (*MinUtility + (1-r)*[MaxUtility-MinUtility]*) with r∈ [0, 1]. A random machine is selected from the eligible set to assign the process to it. The solution is declared infeasible if there is no machine to host it and removed from the initial solutions. Then at the Global step set of decent solutions spread over the search space. Therefore, the algorithm exhibits complex dependencies on constraints and services. Thus, the assignment of the resources is not straightforward and needs the incorporation of efficient resource utilization and SLA violation. In this way, there is a profound effect on the datacenter energy consumption.

An ACS-based VM Consolidation (ACS-VMC) method is proposed by Farahnakian, Ashraf [94] to consolidate VMs for Green Cloud Computing. The authors formulated an energy-efficient VM consolidation as a multi-objective optimization problem to optimize three conflicting objectives. That is, minimizing energy consumption, VM migrations, and avoiding SLA violations. A multi-dimensional resource utilization of PM (CPU, memory and network components) is used together with distributed multi-agent system architecture for dynamic VM consolidation as shown in Figure 8. The figure consists of two types of agents: local and global. A Local agent resides in a PM to solve the PM status detection sub-problem by observing the current resource utilization of the PM while the Global agent acts as a supervisor and optimizes the VM placement by using the proposed ACS-VMC algorithm.

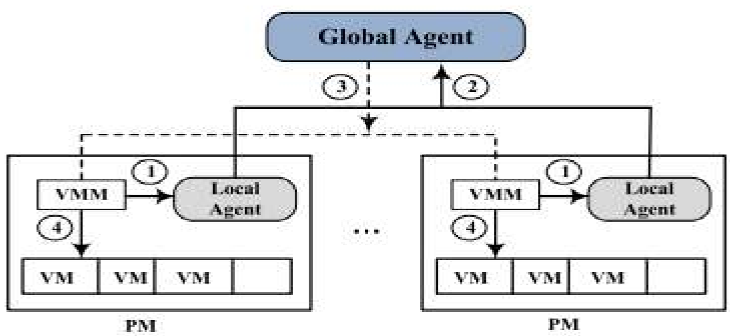

Fig 9. the system Architecture [94]

The ACS-VMC creates a set of tuples T, where each tuple, $t \in T$, consists of three elements: source PM p_{so}, VM to be migrated v, and destination PM p_{de}, as shown in Eq 12.

$t = (p_{so}, v, p_{de})$ (12)

ACS-VMC employs the Travelling Salesman Problem (TSP) methods to reduce the computational time of PMs AND VMs consolidation that depend on the number of tuples (T). Besides, it applies two constraints when making the set of tuples, T, as stated in Eq 13 and 14.

$p_{so} \in \hat{P}_{over} \vee p_{so} \in P_{over} \vee p_{so} \in P_{under}$ (13)

$$p_{de} \notin \quad P_{over} \wedge p_{de} \notin \hat{P}_{over} \tag{14}$$

These constraints check the future prediction of the over and under-loaded PMs used as a source PM denoted as p_{so} . While P_{over} represents the overloaded PM, and P_{under} is the under-loaded PM at the destination. The computational time is reduced without compromising the quality of the results. And the output of the VMs consolidation becomes the VM migration. The metrics of energy consumption and the SLA violation are combined because of their relationship regarding performance, as shown in Eq 15.

$$ESV = EC\ X\ SLAV \tag{15}$$

where ESV is the, EC is the energy consumption, and $SLAV$ is the service level agreement violation. The method aims to reduce datacenter energy consumption by consolidating VMs into a reduced number of active PMs. But it did not consider the network component of the datacenter. Moreover, it exhibits low workload utilization due to the nature and characteristics of ACS-VMC on the workload traces. In the light of this, the ACS-VMC may not be energy-efficient.

Sait, Bala [95] proposed a Multi-Objective CSO Algorithm (MO-CSOA) to minimize the number of PMs used for reducing energy consumption. The technique starts by setting initial parameters such as a fraction the population in the bottom nests denoted as *Pa,* and the maximum number of iterations, *MAXiter*. The results in partitioning of the population into top and bottom nests. The initial population of size, *S, is* generated by first of all obtaining *S* random permutations of the VM requests. To obtain *S* different placement solutions to serve as the initial population at the end. In each next step, a fitness function is found by computing the average fitness of placed VMs based on the fitness of a VM. Similarly, the new nest generated is made to replace the old bottom nest. The population is then partitioned again into top and bottom nests using perturb_1 and perturb_2 that receive nest as input. The fitness of nest is obtained by applying fuzzy evaluation method using perturb_1 and perturb_2 to receive nest as an input. Then, the power consumption and resource wastage objectives are combined into one objective function with upper and lower bound states defined for server power and resource wastage. The technique simultaneously optimizes the power consumption and resource wastage of the datacenter by modifying the tuning parameters of the algorithm. However, the algorithm supports static VM placement and requires SLA violation consideration. This implies that MO-CSOA is not reliable regarding resource utilization and energy-efficiency.

Marotta and Avallone [96] have proposed a combined Mixed Integer Linear Programming (MILP) and Simulated Annealing (SA) for energy consumption reduction in a datacenter. An initialization phase starts with the SA in which an initial feasible solution for the problem is constructed. Then, different solutions are explored by performing some iterations. At the end of each iteration, the temperature value is updated. The algorithm ends when the temperature value goes below a given threshold. A single iteration involves generating many solutions, each of which is obtained by perturbing the current solution through a random move. Each new solution is accepted completely if it is allowed to improve the objective function value, or with a certain probability, which is a function of the temperature and the difference between the two objectives. The SA saves energy and improves datacenter performance by putting idle network components into sleep modes. It is unable to handle longer computation time, and consolidation decision does not consider traffic among the VMs. Therefore, the SA method may not be as energy efficient as claimed.

VM Consolidation in cloud datacenters using ACO metaheuristics (VMC-ACO) is proposed by Ferdaus, Murshed [97]. VMC-ACO employs a new version of ACO algorithm that uses vector algebra. It assumes a homogeneous environment and considers CPU, network I/O and memory as significant resources. The paper explores the bin packing problem and models physical machines (PMs) as bins and VMs as items to pack into the bins. The objective function, *f,* is formulated as a single minimization function on *y* as shown in Eq 16.

$$min\ f(y) = \sum_{i=0}^{n} y_i \tag{16}$$

To capture the degree of imbalance in current resource utilization of a PM, the Resource Imbalance Vector (RIV) is applied which is computed as vector difference as expressed in Eq 17.

$$magRIV = \sqrt{(C-H)^2 + (M-H)^2 + (I-H)^2} \tag{17}$$

Likewise, resource wastage is modeled as the summation of the remaining resources of each resource in Equ 12 and the power consumption by the PM p is modeled as a linear function of CPU utilization as shown in Eq 18 and 19.

$$Wastage_p = \sum_{r \in R} 1 - U_p^r \tag{18}$$

$$E(p) = \begin{cases} E_{idle} + (E_{full} - E_{idle}) + U_p^{CPU} & if\ U_p^{CPU}\ > 0 \\ 0 & otherwise \end{cases} \tag{19}$$

VMC-ACO uses a recent version of ACO and considers each VM-to-PM assignment as a solution component. Pheromone levels are associated to all VM-to-PM assignments representing the desirability of assigning a VM to a PM (Eq 18 & 19). The VMC-ACO has the advantage of balancing the utilization of infrastructure resources and therefore, reduces the energy consumption of the datacenter. However, VMC-ACO has not considered the resource utilization of network components during the VM placement decision. The technique has an adverse impact on the overall resource allocation of the datacenter, and thus limits its energy efficiency.

A modified PSO (MPSO) to consolidate VMs to avoid falling into a track of local optima is proposed by Li, Zhu [98]. MPSO takes into consideration the influence of multi-resource scheduling and allocation of VM on Cloud environment. It focuses only on the CPU and disk to measure the resource utilization and energy consumption but leaves out other components of the server and network. A multi-resource double threshold method is used to trigger the VMs migration. Then the migration policy is applied to select the VMs to migrate. Then a modified PSO method is introduced into VMs reallocation algorithm to reduce the energy consumption of the whole system. Thus, although MPSO reduces energy consumption, the results it produces may seem inadequate since memory and networks components are not considered.

Gabaldon, Guirado [99] have proposed a PSO-AE (Energy-Aware) technique that uses computational resources of a real workload traces to determine the task allocation to available resources with minimum energy consumption. In PSO-AE, each particle consists of two parts. Figure 9 shows how the first part represents the order in which the jobs must be executed in the system. The second part represents a list of forbidden nodes for each job. The list is implemented as a real number in the range [0, 1] for each job and cluster.

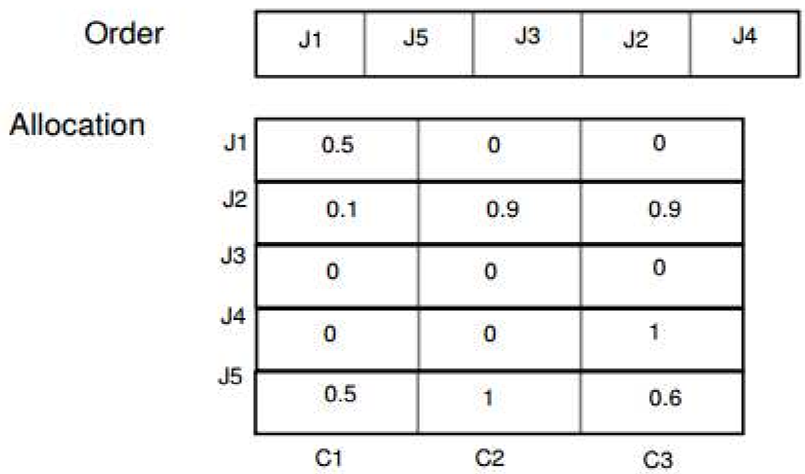

Fig 9. PSO Representation [99]

PSO-AE is used on web workload applications and changes the resource allocation requirement, leading to the use of dynamic resource reconfiguration. It proofs to be effective in reducing the energy consumption of datacenters. However, it shows low sensitivity to workloads and SLA violation for different applications. This makes it less energy-efficient.

3.5. Comparative assessment of energy-efficient techniques

Table 2 shows comparisons of the Nature-Inspired algorithms that have been explained in the above sections and present significant remarks about their achievements.

1. Algorithm

It can be observed from the Table 2 that some of the algorithms are hybrids of others algorithms. An example of such hybrid types includes DVFS-MODPSO, by Yassa, Chelouah [38], Javanmardi, Shojafar [90], Shojafar, Javanmardi [91], and Hybrid ACO & CS, by Moganarangan et al. (2016). We note that the hybrid approach is usually adopted as a way of achieving good results regarding energy-efficiency and performance. We further note that most of the algorithms used under this approach are Evolutionary Intelligence, Swarm Intelligence, and Bio Intelligence. Clearly, the use of these algorithms produce good outputs. Nevertheless there is still the need for more explorations into other Nature-Inspired algorithms that can enhance the current ones and achieve better results.

2. Technique

Table 2 shows that the energy-efficient techniques used in Cloud datacenters are mainly three; Virtualization, Consolidation, and Energy-aware. These techniques are used to identify resources that are considered to be fully utilized and consume the larger energy. It also noted that the Virtualization and Consolidation techniques are used the more. This is due to the flexibility and universal feature of easy implementation which make them more popular techniques than others.

3. Parameters

The parameters used by the techniques to determine their performance are shown in Table 2. The choice of the parameters depends on the Cloud user perspective or the Cloud, service provider. We observed that the provider is concerned more with efficient resource utilization while the users are focused on application performance. In that regards, the user's parameters include makespan, response time, execution time, fairness, turnaround time and tardiness, and the provider parameters include energy consumption reduction, resource utilization, VM migration, workload, throughput, budget and other dependency constraints (death line, reliability and priority constraint). All these parameters are put together in the objective function of the particular Nature-Inspired algorithms while scheduling on the Cloud resources.

4. Benchmarking

The benchmarking (i.e., comparison method) shows the extent of improvement of the proposed algorithms regarding energy reduction. Some of the techniques are compared with the traditional heuristics algorithms such as Round Robbin, HEFT Heuristics, First Fit Decreasing and Modified First Fit Decreasing Algorithms whereas others are compared with the current classical Nature-Inspired Algorithms such as ACO, PSO, GA, and CSO. It is observed that energy-aware VM allocation and placement, dynamic workload allocation or migration and other policies are used to measure energy-efficiency of the proposed algorithms. These techniques are implemented in different simulation environment with real data traces and further validated through empirical analysis and hypothesis test to show their performances in reducing the datacenter energy consumption.

5. Advantage

The Nature-Inspired algorithms are considered to be robust and advantageous over Heuristics or Non-Nature-Inspired algorithms, regarding speed, optimal solutions and computational complexity. Table 2 shows that the proposed algorithms and techniques have various advantages ranging from faster convergence, excellent performance, dynamicity, feasible assignment, improvement of energy-efficiency, and resource utilization.

6. Limitation

From the table, it can also be observed that the algorithms contain many drawbacks. Among them include SLA violations, lack of computational complexity and use of a single resource, task dependency, poor reliability, in-efficient resource utilization, slow convergence /response time and the neglect of other energy consuming components such as network resource and I/O resources. However, the central limitations of each of the techniques are also summarized in the following paragraphs:

Virtualization technique does not guarantee a reduction in energy consumption due to the high communication overhead and considers single resource (Processor) utilization at a time. More so, the technique proposed by Wang, Wang [80], Wang, Wang [81], Ramezani, Lu [82] and Yao, Ding [83] take longer times to reach the no dominated solution with poor migration concept to reduce the datacenter underutilization, and energy consumption.

Consolidation technique has overcome most of the limitation with the help of virtualization technique as proposed by Ferdaus, Murshed [97], Marotta and Avallone [96] and Sait, Bala [95]. Still, the technique also suffered from low sensitivity on the workload traces that lead to low workload utilization and did not explore in detail the network communication overhead during the VM migration. Furthermore, the consolidation technique raises the issue of consolidating the PM resources with the network resources together as VMs can move from one PM to another, and it has been a neglected component (network) in this domain.

Energy-Aware technique such as Mezmaz, Melab [67], Malakooti, Sheikh [68], Raju, Amudhavel [69] and Kansal and Chana [73] focuses more on VM migration and powering on/off PM to minimize the total energy consumed by the PMs that also result into some problems such as high convergence, lack of performance guarantee, SLA violation and as well under-utilization of resources.

7. Energy Efficiency

We note that not all the techniques are energy-efficient as claimed by their authors. The reasons for their low performance are due to the limitations and the parameters used in their problem formulations and implementations. Therefore, we categorized them into high, medium and low energy-efficient performing algorithms.

3.5.1. Observation summary

Much of the results are preliminary, in the sense that many of the current strategies are evaluated via simulation. The implementation of a simulation framework constitutes a significant contribution because the discrete-event simulation is usually chosen as a first step toward creating Cloud Computing environments due to the ability to guarantee repeatable conditions for such a set of experiments. For example, Kessaci, Mezmaz [29] have used the Grid environment for implementation and map-reduce framework has been used by Raju, Amudhavel [69] while the remaining technique such as [82, 95] have used the Cloudsim environment. Almost all the proposed technique that is implemented using single datacenter architecture and very few techniques use multi-datacenter architecture as proposed by Phan, Suzuki [85] and Ramezani, Lu

[82] in their scheduling optimization model. Due to the complexity of data center infrastructures in deploying and configuring resources in power- and energy-efficient manner as well as fulfilling several constraints at the same time. The constraint ranges from makespam, resource utilization, runtime, execution time, migration time, energy consumption, etc., and as a result, several important problems remain open. For example, despite the fact that consolidation of VMs contributes to improving energy efficiency in the data center, limitations in virtualization technology can lead to interference among VM instances in a physical server, which may degrade the performance of applications running on those VMs. Most papers also provide some evaluation of the algorithms they propose. In most cases, that assessment is done empirically, but there are also some examples of rigorous mathematical analysis as presented in Ferdaus *et al.* (2014), Farahnakian, Ashraf [94] and Sharma and Reddy (2015) that are being used in energy efficient scheduling techniques. Likewise, the virtualization and consolidation techniques focus on resolving the resource management and energy consumption problem based on the three decision questions namely (i) when to migrate a particular VM (ii) where to migrate the VM and (iii) which PM to switch off to determine the best placement resources that result in an optimal resource and energy consumption.

Table 2. Comparison of Nature-Inspired energy-efficient techniques

Algorithm	**Energy Efficient Technique**	**Approach**	**Scheduling Method**	**Problem Formulation**	**System Resource**	**Measured Parameter**	**Benchmark**	**Advantage**	**Limitation**	**Energy-Efficiency**
H- DVFS-MODPSO Yassa et al. (2013)	Energy-Aware	MOO	Static/dynamic	Work Flow Scheduling	PM	Execution Time, Cost& EC	HEFT Heuristic	improves energy consumption & makespam	Not implemented, not reliable	N/A
MOCell, NSGA-II and IBEA algorithms Guzek et al. (2014)		MOO	Static/dynamic	Task Scheduling	CPU	EC & Makespan	HEFT Algorithm	Provide accurate solution for the addressed problem that converge to good solutions	Dependent on task & processor number	Low
EA-ACO Feller et al. (2011)		SOO	Static	Resource Allocation	CPU, RAM, DISK & Network	RU & EC	First-Fit Decreasing (FFD)	Achieve higher energy saving & resource utilization	Does not support heterogeneity	Medium
EAVM-ACO Liu et al. (2014)		SOO	Dynamic	Resource Allocation	CPU & Memory	mber of VM and Servers	First-Fit Decreasing (FFD) Algorithm	Minimize energy consumption & resource wastage	High convergence time	Low
FOA Kansal and Chana (2016)		SOO		Resource Allocation		EC, RU & Migration Time	ACO-based & FFD-based Algorithms	Maintained good energy-efficiency & performance	No performance guarantee	High
PreAntPolicy Duan et al. (2016)		SOO	Dynamic	Resource Allocation	CPU & PM	EC, CPU Utilization & CPU Load	First-Fit, Round-Robin & MM	Minimize energy consumption	under-utilization of resources	Medium
EMOA Phan et al. (2012)	Virtualization	MOO	Dynamic/Static	Resource Allocation	PM	Renewable EC, Cooling & User-to-Service Distance	Static & Dynamic Placement Algorithms	Improves renewable energy consumption	SLA violation has not been consider, slow response time	High
EOA Pascual et al. (2015)		MOO	Dynamic	Resource Allocation	PM & Network	EC, ET & RU	NSGA-2, SPEA-2 & Hype	Reduces power & faster processes of request	High Communication overhead	Medium

OL-PICEA-g Lei et al. (2016)		MOO	Dynamic	Task Scheduling	PM	EC, Makespan, Utilization of Renewable Energy & Task Satisfaction	PICEA-g Algorithm	reduces makespan & energy consumption	It does not handle parallel task scheduling	Low
Hybrid ACO & CS Moganarangan et al. (2016)		SOO	Dynamic	Task Scheduling	PM	EC & Makespan	ACO	Substantially reduce energy consumption	Consider the energy consumption of processors only	Medium
GAHO-ILP Rocha and Cardozo, 2014)		SOO	Dynamic	Resource Allocation	PM & Network	EC, PACKET LOSS & SPEED	OSPF	Trade-off between server energy consumption & network	Takes longer time to reach the non-dominated solutions	Medium
DVFS-GA Sharma and Reddy (2015)		SOO	Static	Resource Allocation	PM	EC & RU	Multi-objective Genetic Algorithm	Save energy with 0% SLA Violation	Lack VM migration concept	Low
Hybrid ACO & CS Moganarangan et al. (2016)		SOO	Dynamic	Task Scheduling	PM	EC & Makespan	ACO	Substantially reduce energy consumption	Consider the energy consumption of processors only	Medium
GeNePi Saber et al. (2014)	Consolidation	MOO	Dynamic	Resource Allocation	RAM & CPU	Reliability, Migration Cost & EC	Firs Fit (ff), Balancing Bin (BB) & Random Fit (RF)	Finds non-dominated solution easily	SLA Violation is not consider	Low
ACS-VMC Farahnakian et al. (2015)		MOO	Dynamic	Resource Allocation	PM & Network	SLAV, EC & VM Migration	MAD, IQR, LR & THR Heuristics	Reduces Ec & maintained QoS	Low workload utilization level	High
CSO-A Sait et al. (2016)		MOO	Dynamic	Resource Allocation	CPU	Convergence Rate, Reliability, EC& RU	GGA, RGGA, ILL& IFFD	It uses dynamic placement of VMs	Not reliable	Low
PSO-AE Gabaldon et al. (2016)		SOO	Dynamic	Resource Scheduling	PM	EC	JPR-E, FCFS, Min-Min & HILL	Faster convergence with lowest energy consumption	Low sensitivity to workloads	Low
MPSO Li et al. (2016)		SOO	Dynamic	Resource Allocation	CPU & Disk	EC, Migration, Load & Balancing	Modified Best Fit Decrease	Indicate better energy-efficiency and reduces VM migration	Lack SLA violation, consider only CPU & Disk	Low

SA- MILP Marotta and Avallone (2015)	SOO	Static	Resource Allocation	CPU & RAM	EC, Migration, RU & Makespan	First Fit Decreasing (FFD) & Sercon	Can find feasible assignment easily	The consolidation decision does not consider traffic among the VMs	Medium
VMC-ACO Ferdaus et al. (2014)	SOO	Static	Resource Allocation	CPU, Memory & I/O	Resource Wastage & Runtime	Max-Min Ant System, VectorGreedy Algorithm, Modified FFD-Volum & FFD-L1Norm	Applicable in large virtualize datacenters	Did not consider network utilization and live VM migration	Low

4. Future research direction

In spite of the abundance of literature available on this topic, this study has identified certain aspects that have potentials for further exploration. This section highlights some of these areas.

4.1. New design environment for benchmarking Nature-Inspired technique

Energy efficiency techniques is an issue of interest for investigation. In most of the analysis of results of these techniques, there is a need for stability or consistency in performance; since most Nature-Inspired algorithms are heuristic. Therefore, balance issues of these algorithms in respect to development environment needs further exploration. More so, most of the techniques are implemented using simulations environment and are evaluated on different workloads. However, the success of a particular Nature-Inspired technique in achieving optimal energy efficiency is dependent on its design environment (i.e., encoding scheme, operators, set of parameters, etc.). So for a given complex problem, the design choices should be theoretically analyzed before the simulation and implementation. There is a need for improvement in the way these algorithms are implemented. Therefore, to solve any energy consumption problem in Cloud Computing domain results from tests in real-case scenarios are needed to assess the performance better and eventually mitigate some limitations of the proposed models and algorithms. In most of the analysis of results of energy efficiency usage, there is a need for stability or consistency in performance of the tools used to measure the resources at the IaaS level.

4.2. Dual service level agreement and security-aware scheduling

Various Nature-Inspired energy-efficient techniques have been presented in the previous sections, but the ratio of SLA violations and security is still overwhelming. Research can be done to incorporate Dual SLA with customers while performing energy-aware scheduling, as most of the techniques compromise performance in trying to reduce the datacenter energy consumption with efficient resource allocation. For example, Khoshkholghi, Derahman [100] and Singh *et al.* (2017) proposed a new technique that measures energy consumption with SLA violation Time per Active Host (SLATAH) and Energy SLA violation (ESV) to guarantee system performance under the SLA constraint. Providers can also negotiate dual SLA with the user, due to the transition delay of PMs from inactive to active mode. The IaaS model has taken the accessibility issue and pushed other security and privacy issues to Cloud users [101]. Figure 10 shows the concept of the dual SLA and security-aware scheduling as the future research direction. The second SLA is optional and can be opted by the Cloud user only when he wants to be in ‘‘Green mode”. ‘‘Green mode” means the primary objective is energy optimization, and the performance may be somewhat compromised but within acceptable limits and the cost savings, thus achieved can be used to benefit the customer also. Security and privacy aware scheduling is another area that needs to be explored using Nature-Inspired techniques. The proposed works concentrate on administering suitable VM allocation taking into consideration of SLA. The current datacenters suffer from DoS/DDoS attack due to susceptibility which results in higher energy consumption and CO_2 emission [102]. Investigations are required to perform scheduling in a way that it protects the sensitive and private information associated with users. This type of schedule is important when the scheduled jobs carry confidential and personal information about various subjects in a given context.

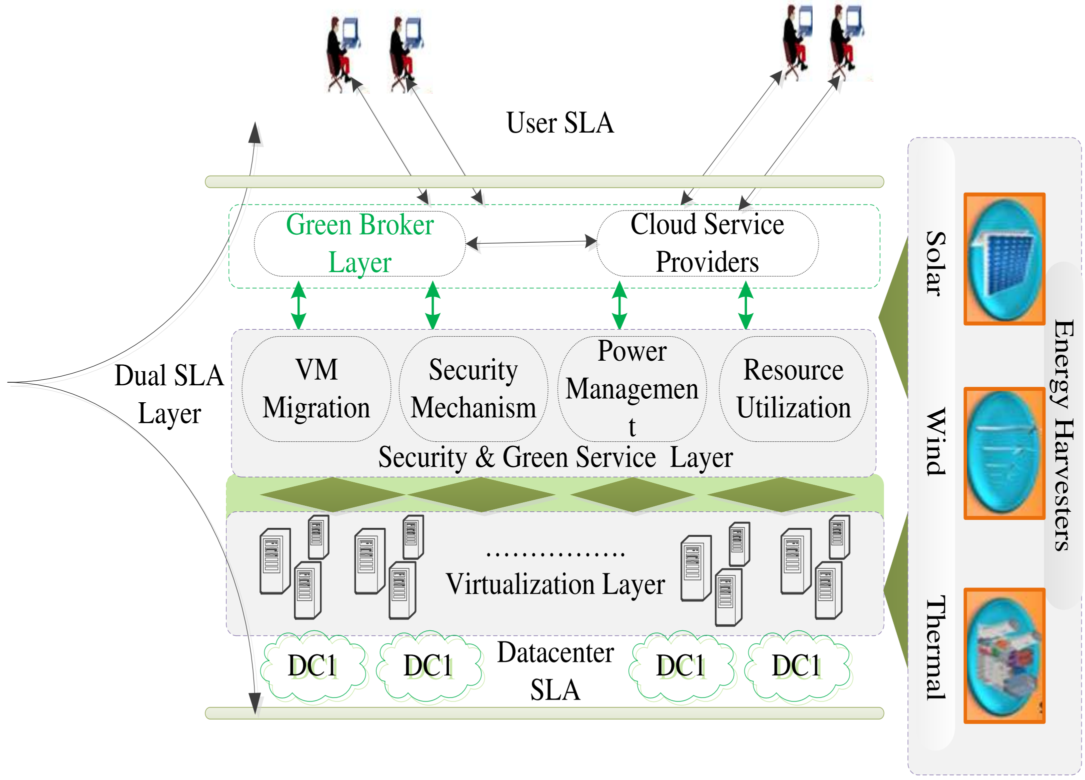

Fig 11. Proposed dual service level agreement and security-aware scheduling

4.3. Energy-Aware workload constraint in datacenters

Workload refers to the distribution of running applications, storages, backups, and databases across PMs at different locations of the Cloud datacenter for best performance delivery. However, due to the flexibility characteristics of datacenter deployment, different applications require different resources such as CPU, memory, storage and network bandwidth with each resource having a constraint regarding response time, throughput, execution time, energy consumption and resource utilization. Research has shown that integrating CPU-intensive workload and network-intensive workload incurs the least resource contention, thus improving the combined performance [87]. Although existing research on consolidation techniques try to address this problem by considering the resource constraint, for example, the machine reassignment algorithm for datacenters called *GeNePi* was proposed by Saber, Ventresque [93] while Farahnakian, Ashraf [94] consolidate datacenter workloads for Green Cloud Computing. The workloads consume a significant amount of datacenter resources to mitigate the risk associated with configuration set up, VM migration as well as resource isolation. Research is encouraged to investigate the types of workload which can be efficiently connected to a PM while performing consolidation of VMs for energy optimization and resource utilization. This will significantly improve resource provisioning, as well as consistency in deploying applications, in addition to energy saving in the datacenter environment.

4.4. Resource management for sustainable multi-tenant and reliable Cloud datacenters

Cloud datacenter resource allocation, provisioning, sharing, and sustainability are interesting topics in Cloud Computing domain. They are used in minimizing the capital expenditure of Cloud service providers. Sustainability in Cloud datacenter influences its economic and environmental impact to a large extent. However, the overall sustainable architecture of Cloud Computing datacenter is subject to many issues like the assurance of QoS, service reusability, energy efficiency, resource management and so on. In the existing literature, there is little discussion about sustainable and reliable datacenters such as the work presented by Farahnakian, Ashraf [94] and Faruk, Ruttik [103]. More so, available datacenter resources can be virtualized and allocated to multiple users across different datacenters as proposed by Kansal and Chana [73]. Furthermore, multi-tenant support in datacenter resource management as well as scheduling the computational tasks and their respective workloads according to their energy source, request capacity, and QoS requirements have not been investigated in great detail. Figure 11 shows the framework of the proposed resource management for sustainable multi-tenant datacenters.

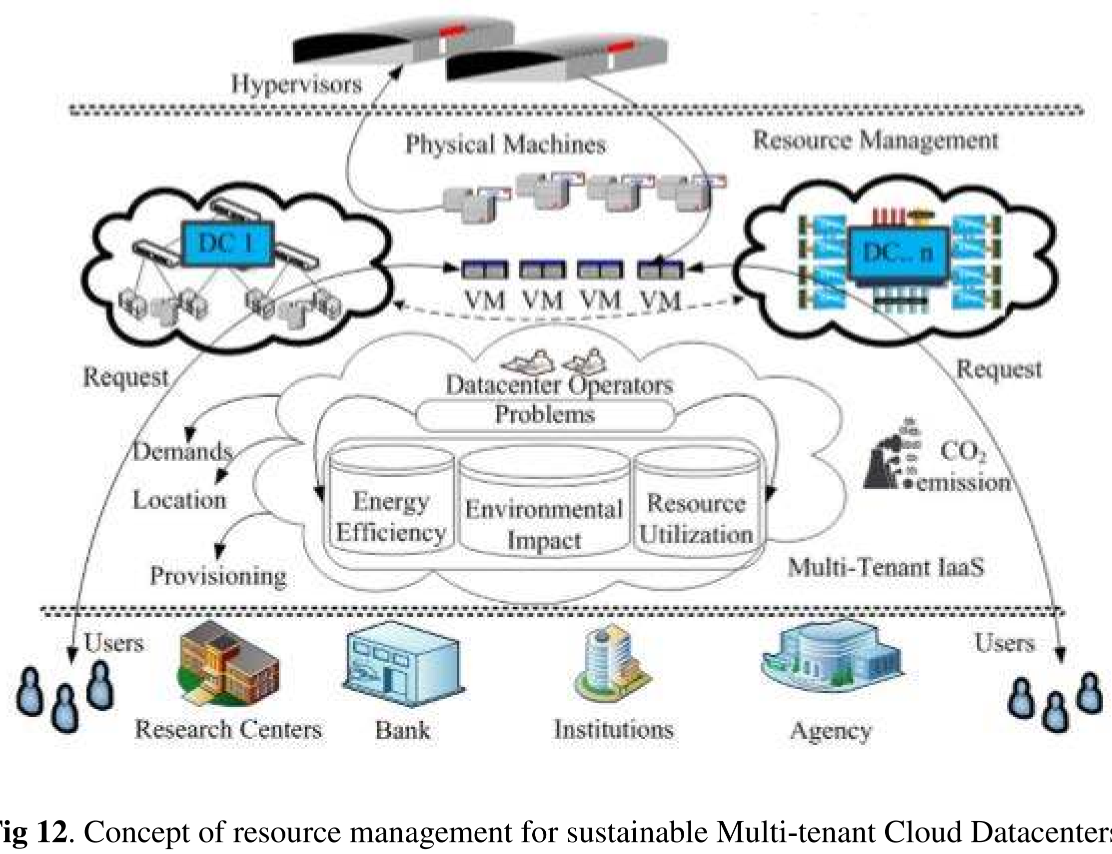

Fig 12. Concept of resource management for sustainable Multi-tenant Cloud Datacenters

Therefore, future research can be conducted targeting this limitation of existing literature that can lead to a reduction of resource overhead when managing multiple datacenters and carbon emission; thereby, minimizing the sources of the environment pollution.

1. Integrating solution of Cloud datacenters with Fog computing and Big Data

It would be interesting and potentially rewarding to investigate the role of energy consumption as a whole in the growth of Cloud Computing and its resource scheduling in the presence of Big Data and Fog Computing through the Internet. An industrial example is that of automation systems via the Internet of Things (IoT) such as sensors, actuators, and smart cities have been developed fast with potentially massive market demands in Cloud Computing and Big Data for industries. Unlike the earliest attempt by the Grid and Distributed computing, Cloud Computing models offer automation functions as services from a dynamic infrastructure that supports operational complexity and resource heterogeneity. For example, there is research going on that integrates Fog with Cloud and Big data to reduce energy consumption and improve analytic solution [104-106]. The integrated energy-efficient scheduling in Cloud, Fog and Big data stream environments to minimize response time and as well energy consumption is shown in Figure 12.

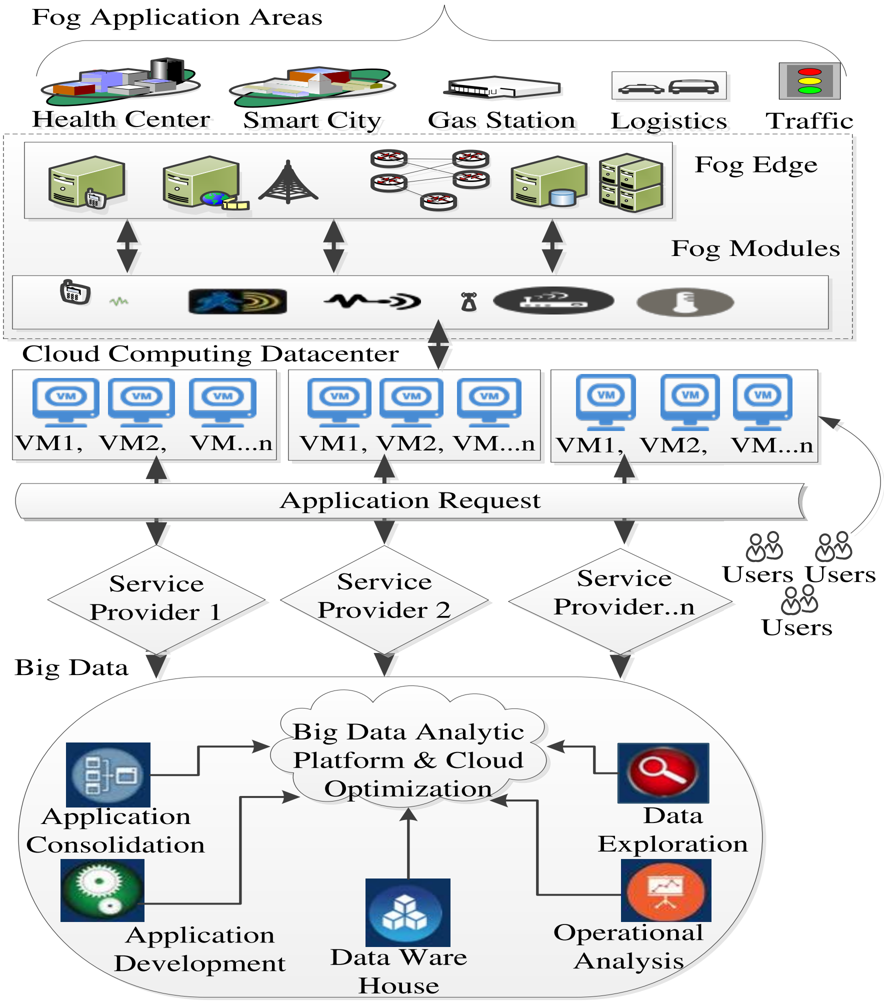

Fig 13. Integrated energy-efficient resource scheduling in Cloud Fog and Big data

Unlike the earliest attempt by the Grid and Distributed Computing, Cloud Computing models offer automation functions as services from a dynamic infrastructure that supports operational complexity and resource heterogeneity. Adopting this concept will open new areas of research and collaboration that will bring solutions not only to single data operation but bringing in all devices that generate data to the maximum utilization.

5. Conclusion

This paper reviewed Nature-Inspired algorithms, techniques, tools, and methods commonly employed to mitigate energy consumption in Cloud computing datacenters. The techniques studied were analyzed based on their goals, methods, strengths, and limitations. Furthermore, the studied algorithms were examined based on their features to establish their respective energy efficiency levels. Even though the techniques work well at different levels of the IaaS, the analysis reveals many limitations in their techniques. Most existing scheduling techniques require large resources that consume high energy of the Cloud datacenters. The resource allocation strategies adopted by these techniques are slow in convergence, leading to high energy consumption. The network component which consumes a significant amount of energy of the datacenters is conspicuously neglected in the existing techniques. In some scenarios, the datacenters operate in silos (i.e., single datacenters instead of multi datacenters), this results in underutilization of the datacenter infrastructure and thus, wasting away energy. Analysis of the techniques further revealed that in attempts to reduce energy consumption, SLAs are violated. Given the limitations of the existing techniques, there is the need to develop better techniques that can efficiently and proactively improve the energy efficiency as well as resource utilization of Cloud datacenters. Such techniques must be designed to control resource and energy wastage without ignoring energy consumption made by networking components, and input-output devices. In this way, the global carbon emission can be reduced substantially, ensuring environmental sustainability and decreasing adverse effects on human life.

Conflict of Interest

Authors declare that there is no conflict of interest regarding the manuscript.

Reference

1. Foster, I., et al. *Cloud computing and grid computing 360-degree compared*. in *2008 Grid Computing Environments Workshop*. 2008. Ieee.
2. Beloglazov, A., J. Abawajy, and R. Buyya, *Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing.* Future generation computer systems, 2012. **28**(5): p. 755-768.
3. Jiang, D., Z. Xu, and Z. Lv, *A multicast delivery approach with minimum energy consumption for wireless multi-hop networks.* Telecommunication systems, 2016. **62**(4): p. 771-782.
4. Jiang, D., et al., *Energy-efficient multi-constraint routing algorithm with load balancing for smart city applications.* IEEE Internet of Things Journal, 2016. **3**(6): p. 1437-1447.
5. Buyya, R., C.S. Yeo, and S. Venugopal. *Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities*. in *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on*. 2008. Ieee.
6. Beloglazov, A. and R. Buyya. *Energy efficient resource management in virtualized cloud data centers*. in *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. 2010. IEEE Computer Society.
7. Yeluri, R. and E. Castro-Leon, *Cloud Computing Basics*, in *Building the Infrastructure for Cloud Security*. 2014, Springer. p. 1-17.

8. Prasanth, A., et al., *Cloud Computing: A Survey of Associated Services.* Book Chapter of Cloud Computing: Reviews, Surveys, Tools, Techniques and Applications-An Open-Access eBook published by HCTL Open, 2015.
9. ENERGY, S., *Report to congress on server and data center energy efficiency public law 109-431.* Public law, 2007. **109**: p. 431.
10. Dou, H., et al., *A two-time-scale load balancing framework for minimizing electricity bills of internet data centers.* Personal and Ubiquitous Computing, 2016. **20**(5): p. 681-693.
11. Fister Jr, I., et al., *A brief review of nature-inspired algorithms for optimization.* arXiv preprint arXiv:1307.4186, 2013.
12. Mishra, K., S. Tiwari, and A. Misra. *A bio inspired algorithm for solving optimization problems*. in *Computer and Communication Technology (ICCCT), 2011 2nd International Conference on*. 2011. IEEE.
13. Usman, M.J., A.S. Ismail, and H. Chizari, *THE COLLOQUIUM.*
14. Beloglazov, A., et al., *A taxonomy and survey of energy-efficient data centers and cloud computing systems.* Advances in computers, 2011. **82**(2): p. 47-111.
15. Jing, S.-Y., et al., *State-of-the-art research study for green cloud computing.* The Journal of Supercomputing, 2013. **65**(1): p. 445-468.
16. Kaur, T. and I. Chana, *Energy Efficiency Techniques in Cloud Computing: A Survey and Taxonomy.* ACM Computing Surveys (CSUR), 2015. **48**(2): p. 22.
17. Madni, S.H.H., M.S.A. Latiff, and Y. Coulibaly, *An appraisal of meta-heuristic resource allocation techniques for IaaS cloud.* Indian Journal of Science and Technology, 2016. **9**(4).
18. Madni, S.H.H., M.S.A. Latiff, and Y. Coulibaly, *Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities.* Journal of Network and Computer Applications, 2016. **68**: p. 173-200.
19. Kalra, M. and S. Singh, *A review of metaheuristic scheduling techniques in cloud computing.* Egyptian informatics journal, 2015. **16**(3): p. 275-295.
20. Hameed, A., et al., *A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems.* Computing, 2014: p. 1-24.
21. Kołodziej, J., S.U. Khan, and A.Y. Zomaya, *A taxonomy of evolutionary inspired solutions for energy management in green computing: problems and resolution methods*, in *Advances in Intelligent Modelling and Simulation*. 2012, Springer. p. 215-233.
22. Goldberg, D.E., *Genetic algorithms in search, optimization, and machine learning, 1989.* Reading: Addison-Wesley, 1989.
23. Engelbrecht, A.P., *Fundamentals of computational swarm intelligence*. 2006: John Wiley & Sons.
24. Knauth, T., *Energy Efficient Cloud Computing: Techniques and Tools*. 2014, Saechsische Landesbibliothek-Staats-und Universitaetsbibliothek Dresden.
25. Brill, K.G., *Data center energy efficiency and productivity*. 2007, Santa Fe, NM: The Uptime Institute. www. uptimeinstitute. org/symp_pdf/(TUI3004C) DataCenterEnergyEfficiency. pdf.
26. Koomey, J.G., *Worldwide electricity used in data centers.* Environmental Research Letters, 2008. **3**(3): p. 034008.
27. Jiang, D., et al., *QoS constraints-based energy-efficient model in cloud computing networks for multimedia clinical issues.* Multimedia Tools and Applications, 2016. **75**(22): p. 14307-14328.
28. Snowdon, D.C., S. Ruocco, and G. Heiser, *Power management and dynamic voltage scaling: Myths and facts.* 2005.
29. Kessaci, Y., et al., *Parallel evolutionary algorithms for energy aware scheduling*, in *Intelligent Decision Systems in Large-Scale Distributed Environments*. 2011, Springer. p. 75-100.
30. Kliazovich, D., P. Bouvry, and S.U. Khan, *DENS: data center energy-efficient network-aware scheduling.* Cluster computing, 2013. **16**(1): p. 65-75.
31. Meisner, D., B.T. Gold, and T.F. Wenisch. *PowerNap: eliminating server idle power*. in *ACM Sigplan Notices*. 2009. ACM.
32. Deng, Q., et al., *Memscale: active low-power modes for main memory.* ACM SIGARCH Computer Architecture News, 2011. **39**(1): p. 225-238.

33. Sardashti, S. and D.A. Wood. *UniFI: leveraging non-volatile memories for a unified fault tolerance and idle power management technique*. in *Proceedings of the 26th ACM international conference on Supercomputing*. 2012. ACM.
34. Shojafar, M., et al., *Adaptive computing-plus-communication optimization framework for multimedia processing in cloud systems.* IEEE Transactions on Cloud Computing, 2016.
35. Jiang, D., et al., *An optimization-based robust routing algorithm to energy-efficient networks for cloud computing.* 2016.
36. Kim, K.H., A. Beloglazov, and R. Buyya, *Power-aware provisioning of virtual machines for real-time Cloud services.* Concurrency and Computation: Practice and Experience, 2011. **23**(13): p. 1491-1505.
37. Sharma, N.K. and G.R.M. Reddy. *Novel energy efficient virtual machine allocation at data center using Genetic algorithm*. in *Signal Processing, Communication and Networking (ICSCN), 2015 3rd International Conference on*. 2015. IEEE.
38. Yassa, S., et al., *Multi-objective approach for energy-aware workflow scheduling in cloud computing environments.* The Scientific World Journal, 2013. **2013**.
39. Gabrel Torres. *Everything-You-Need-toKnow-About-the-CPU-C-StatesPower-Saving-Modes http://www.hardwaresecrets.com/ december 2015*. Hardware Secrets 2008
40. Snowdon, D.C., et al. *Koala: A platform for OS-level power management*. in *Proceedings of the 4th ACM European conference on Computer systems*. 2009. ACM.
41. Ousterhout, J., et al., *The case for RAMClouds: scalable high-performance storage entirely in DRAM.* ACM SIGOPS Operating Systems Review, 2010. **43**(4): p. 92-105.
42. Koomey, J., *The economics of green DRAM in servers*. 2012, Analytics Press.
43. Hähnel, M., et al. *eBond: energy saving in heterogeneous RAIN*. in *Proceedings of the fourth international conference on Future energy systems*. 2013. ACM.
44. Eom, H., et al., *Evaluation of DRAM Power Consumption in Server Platforms*, in *Ubiquitous Information Technologies and Applications*. 2013, Springer. p. 799-805.
45. Jiang, D., et al., *An optimization-based robust routing algorithm to energy-efficient networks for cloud computing.* Telecommunication Systems, 2016. **63**(1): p. 89-98.
46. Blanquicet, F. and K. Christensen. *Managing energy use in a network with a new SNMP power state MIB*. in *Local Computer Networks, 2008. LCN 2008. 33rd IEEE Conference on*. 2008. IEEE.
47. Michael, A.M. and K. Krieger, *Server power measurement*. 2010, Google Patents.
48. Bianzino, A.P., et al., *A survey of green networking research.* Communications Surveys & Tutorials, IEEE, 2012. **14**(1): p. 3-20.
49. Nie, L., et al. *Traffic matrix prediction and estimation based on deep learning for data center networks*. in *Globecom Workshops (GC Wkshps), 2016 IEEE*. 2016. IEEE.
50. Power, E.N., *Energy logic: reducing data center energy consumption by creating savings that cascade across systems, A White Paper from the Experts in Business-Critical Continuity.* 2008.
51. Cho, J.-K. and S.-H. Shin, *Power and Heat Load of IT Equipment Projections for New Data Center's HVAC System Design.* Korean Journal of Air-Conditioning and Refrigeration Engineering, 2012. **24**(3): p. 212-217.
52. Rivoire, S., et al., *Models and metrics to enable energy-efficiency optimizations.* 2007.
53. Gough, C., I. Steiner, and W. Saunders, *Why Data Center Efficiency Matters*, in *Energy Efficient Servers*. 2015, Springer. p. 1-20.
54. Liu, L., et al. *GreenCloud: a new architecture for green data center*. in *Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session*. 2009. ACM.
55. Belady, C., et al., *Green grid data center power efficiency metrics: PUE and DCIE*. 2008, Technical report, Green Grid.
56. Belady, C., et al., *Carbon usage effectiveness (CUE): a green grid data center sustainability metric.* White Paper, 2010. **32**.
57. Haas, J., et al., *Proxy proposals for measuring data center productivity.* The Green Grid, 2009.

58. Zomaya, A.Y. and Y.C. Lee, *Energy efficient distributed computing systems*. Vol. 88. 2012: John Wiley & Sons.
59. VanGeet, O., W. Lintner, and B. Tschudi, *FEMP best practices guide for energy-efficient data center design.* National Renewable Energy Laboratory, 2011.
60. Newcombe, L., *Data centre energy efficiency metrics.* Data Centre Specialist Group, 2009.
61. Lee, Y.C. and A.Y. Zomaya, *Energy efficient utilization of resources in cloud computing systems.* The Journal of Supercomputing, 2012. **60**(2): p. 268-280.
62. Babukarthik, R., R. Raju, and P. Dhavachelvan. *Energy-aware scheduling using hybrid algorithm for cloud computing*. in *Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on*. 2012. IEEE.
63. Quang-Hung, N., et al. *A genetic algorithm for power-aware virtual machine allocation in private cloud*. in *Information and Communication Technology-EurAsia Conference*. 2013. Springer.
64. Wu, G., et al. *Energy-efficient virtual machine placement in data centers by genetic algorithm*. in *International Conference on Neural Information Processing*. 2012. Springer.
65. Wu, Y., M. Tang, and W. Fraser. *A simulated annealing algorithm for energy efficient virtual machine placement*. in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2012. IEEE.
66. Luo, H., et al. *The dynamic migration model for cloud service resource balancing energy consumption and QoS*. in *Control and Decision Conference (CCDC), 2015 27th Chinese*. 2015. IEEE.
67. Mezmaz, M., et al., *A parallel bi-objective hybrid metaheuristic for energy-aware scheduling for cloud computing systems.* Journal of Parallel and Distributed Computing, 2011. **71**(11): p. 1497-1508.
68. Malakooti, B., et al., *Multi-objective energy aware multiprocessor scheduling using bat intelligence.* Journal of Intelligent Manufacturing, 2013. **24**(4): p. 805-819.
69. Raju, R., et al. *A bio inspired Energy-Aware Multi objective Chiropteran Algorithm (EAMOCA) for hybrid cloud computing environment*. in *Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on*. 2014. IEEE.
70. Feller, E., L. Rilling, and C. Morin. *Energy-aware ant colony based workload placement in clouds*. in *Proceedings of the 2011 IEEE/ACM 12th International Conference on Grid Computing*. 2011. IEEE Computer Society.
71. Liu, X.-F., et al. *Energy aware virtual machine placement scheduling in cloud computing based on ant colony optimization approach*. in *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*. 2014. ACM.
72. Liu, X.-F., et al. *Energy aware virtual machine placement scheduling in cloud computing based on ant colony optimization approach*. in *Proceedings of the 2014 conference on Genetic and evolutionary computation*. 2014. ACM.
73. Kansal, N.J. and I. Chana, *Energy-aware Virtual Machine Migration for Cloud Computing-A Firefly Optimization Approach.* Journal of Grid Computing, 2016. **14**(2): p. 327-345.
74. Duan, H., et al., *Energy-aware scheduling of virtual machines in heterogeneous cloud computing systems.* Future Generation Computer Systems, 2016.
75. A Vouk, M., *Cloud computing–issues, research and implementations.* CIT. Journal of Computing and Information Technology, 2008. **16**(4): p. 235-246.
76. Xu, L., Z. Zeng, and X. Ye. *Multi-objective optimization based virtual resource allocation strategy for cloud computing*. in *Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on*. 2012. IEEE.
77. Song, A., et al. *Multi-objective virtual machine selection for migrating in virtualized data centers*. in *Joint International Conference on Pervasive Computing and the Networked World*. 2012. Springer.
78. Shigeta, S., et al. *Design and implementation of a multi-objective optimization mechanism for virtual machine placement in cloud computing data center*. in *International Conference on Cloud Computing*. 2012. Springer.
79. Gao, Y., et al., *A multi-objective ant colony system algorithm for virtual machine placement in cloud computing.* Journal of Computer and System Sciences, 2013. **79**(8): p. 1230-1242.

80. Wang, X., Y. Wang, and H. Zhu, *Energy-efficient multi-job scheduling model for cloud computing and its genetic algorithm.* Mathematical Problems in Engineering, 2012. **2012**.
81. Wang, X., Y. Wang, and Y. Cui, *A new multi-objective bi-level programming model for energy and locality aware multi-job scheduling in cloud computing.* Future Generation Computer Systems, 2014. **36**: p. 91-101.
82. Ramezani, F., et al., *Evolutionary algorithm-based multi-objective task scheduling optimization model in cloud environments.* World Wide Web, 2015. **18**(6): p. 1737-1757.
83. Yao, G., et al., *Endocrine-based coevolutionary multi-swarm for multi-objective workflow scheduling in a cloud system.* Soft Computing, 2016: p. 1-14.
84. Usman, M.J., et al. *Energy-Efficient virtual machine allocation technique using interior search algorithm for cloud datacenter*. in *Student Project Conference (ICT-ISPC), 2017 6th ICT International*. 2017. IEEE.
85. Phan, D.H., et al. *Evolutionary multiobjective optimization for green clouds*. in *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*. 2012. ACM.
86. Shu, W., W. Wang, and Y. Wang, *A novel energy-efficient resource allocation algorithm based on immune clonal optimization for green cloud computing.* EURASIP Journal on Wireless Communications and Networking, 2014. **2014**(1): p. 1.
87. Pascual, J.A., et al., *Towards a greener cloud infrastructure management using optimized placement policies.* Journal of Grid Computing, 2015. **13**(3): p. 375-389.
88. Lei, H., et al., *A multi-objective co-evolutionary algorithm for energy-efficient scheduling on a green data center.* Computers & Operations Research, 2016. **75**: p. 103-117.
89. Rocha, L.A. and E. Cardozo. *A Hybrid Optimization Model for Green Cloud Computing*. in *Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*. 2014. IEEE Computer Society.
90. Javanmardi, S., et al. *Hybrid job scheduling algorithm for cloud computing environment*. in *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014*. 2014. Springer.
91. Shojafar, M., et al., *FUGE: A joint meta-heuristic approach to cloud job scheduling algorithm using fuzzy theory and a genetic method.* Cluster Computing, 2015. **18**(2): p. 829-844.
92. Moganarangan, N., et al., *A novel algorithm for reducing energy-consumption in cloud computing environment: Web service computing approach.* Journal of King Saud University-Computer and Information Sciences, 2016. **28**(1): p. 55-67.
93. Saber, T., et al. *Genepi: A multi-objective machine reassignment algorithm for data centres*. in *International Workshop on Hybrid Metaheuristics*. 2014. Springer.
94. Farahnakian, F., et al., *Using ant colony system to consolidate vms for green cloud computing.* IEEE Transactions on Services Computing, 2015. **8**(2): p. 187-198.
95. Sait, S.M., A. Bala, and A.H. El-Maleh, *Cuckoo search based resource optimization of datacenters.* Applied Intelligence, 2016. **44**(3): p. 489-506.
96. Marotta, A. and S. Avallone. *A Simulated Annealing Based Approach for Power Efficient Virtual Machines Consolidation*. in *2015 IEEE 8th International Conference on Cloud Computing*. 2015. IEEE.
97. Ferdaus, M.H., et al. *Virtual machine consolidation in cloud data centers using ACO metaheuristic*. in *European Conference on Parallel Processing*. 2014. Springer.
98. Li, H., et al., *Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing.* Computing, 2016. **98**(3): p. 303-317.
99. Gabaldon, E., et al. *Particle Swarm Optimization Scheduling for Energy Saving in Cluster Computing Heterogeneous Environments*. in *Future Internet of Things and Cloud Workshops (FiCloudW), IEEE International Conference on*. 2016. IEEE.
100. Khoshkholghi, M.A., et al., *Energy-Efficient Algorithms for Dynamic Virtual Machine Consolidation in Cloud Data Centers.* IEEE Access, 2017.
101. Kamboj, M. and S. Rana, *Cloud Security and Energy Efficiency.* Advances in Computational Sciences and Technology, 2017. **10**(5): p. 1245-1255.

102. Singh, S., et al., *EH-GC: An Efficient and Secure Architecture of Energy Harvesting Green Cloud Infrastructure.* Sustainability, 2017. **9**(4): p. 673.
103. Faruk, N., et al., *Energy savings through self-backhauling for future heterogeneous networks.* Energy, 2016. **115**: p. 711-721.
104. Masip-Bruin, X., et al., *Foggy clouds and cloudy fogs: a real need for coordinated management of fog-to-cloud computing systems.* IEEE Wireless Communications, 2016. **23**(5): p. 120-128.
105. Stojmenovic, I. *Fog computing: A cloud to the ground support for smart things and machine-to-machine networks*. in *Telecommunication Networks and Applications Conference (ATNAC), 2014 Australasian*. 2014. IEEE.
106. Tang, B., et al. *A hierarchical distributed fog computing architecture for big data analysis in smart cities*. in *Proceedings of the ASE BigData & SocialInformatics 2015*. 2015. ACM.