
Official URL: http://dx.doi.org/10.1007/s42235-019-0030-7
DOI: http://dx.doi.org/10.1007/s42235-019-0030-7
EPrint URI: http://eprints.glos.ac.uk/id/eprint/6797

Disclaimer
The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT.
Virtualization oriented Green Computing in Cloud Datacenter: Flower Pollination Approach

Usman Mohammed Joda1*, Abdul Samad Ismail1, Hassan Chizari2, Gaddafi Abdul-Salaam3, Ali Muhammad Usman4, Abdulsalam Yau Gital5, Omprakash Kaiwartya6, and Ahmed Aliyu7

1. Department of Computer Science, Universiti Teknology Malaysia, 81310 Skudai Johor, Malaysia
2. School of Computing and Technology, Park Campus, University of Gloucestershire
3. Department of Computer Science Kwame Nkrumah University of Science and Technology, Akara Ghana
4. Department of Maths and Computer Federal College of Education Technical Gombe, 072158 Gombe, Nigeria
5. Department of Maths, Abubakar Tafawa Balewa University Bauchi, 81027 Bauchi, Nigeria
6. Department of Computer and Information Technology Northumbria University, Newcastle, NE1 8ST, UK
7. Department of Maths, Bauchi State University Gadau, 81007 Bauchi, Nigeria

*Corresponding Email: umjoda@gmail.com

Abstract

Cloud computing has observed significant interest due to the increasing service demands from organizations offloading computationally intensive tasks to datacenters. Meanwhile, datacenter infrastructure comprises hardware resources consuming a high amount of energy and increasing carbon emissions at a hazardous level. In Cloud datacenter, Virtual Machine (VM) need to be allocated on various Physical Machines (PM) in order to minimize resource wastage and increase energy efficiency. Resource allocation problem is NP-hard, hence finding an exact solution is complicated especially for large-scale datacenters. In this context, this paper proposes an Energy-oriented Flower Pollination Algorithm (E-FPA) for VM allocation in Cloud datacenter environments. FPA is a Natured-Inspired optimization technique used in solving global and numerical optimization problems. A system framework was developed to enable energy-oriented allocation of various VMs on a PM. The allocation uses a strategy namely, Dynamic Switching Probability (DSP). The framework finds near optimal solution quickly and balances the exploration and exploitation of the global and local search. It considers a processor, storage, and memory constraints of a physical machine while prioritizing energy-oriented allocation for a set of virtual machines. Simulations are performed on MultiRecCloudSim utilizing planet workload. It is evident that the E-FPA outperforms state-of-the-art techniques in terms of energy consumption including Genetic Algorithm for Power-Aware (GAPA) by 21.8%, Order of Exchange Migration (OEM) ant colony system by 21.5%, and First Fit Decreasing (FFD) by 24.9%. This implies that, the datacenter performance and environmental sustainability has been improved significantly due reduction in energy consumption and as well carbon emission.

Keywords: Virtualization, Green computing, Cloud, Datacenter, Energy optimization

Copyright © 2018, Jilin University.

1 Introduction

Cloud Computing is a new paradigm that provides computing over the Internet on a pay-per-use basis, and its broader acceptance, coupled with the latest virtualization technologies, contribute to the
establishment of large-scale cloud datacenters to provide the computing services. Viewed from a long-term viewpoint, cloud computing may be perceived as alike to the evolution of the centralized distribution of electricity. The early electricity systems were firstly available on a rather small scale in the form of unconnected networks which slowly moved towards integration and centralization (Lang, 1969). Various services are offered by cloud datacenters to users at different levels.

Dynamic resource allocation consists of automating the allocation or de-allocation of resources in the datacenter without changing the system and or user running applications [1]. Fig 1 is a classification and model diagram of cloud computing showing the various services. Cloud services are categorized into Infrastructure as a Service (IaaS), Software as a service (SaaS), and Platform as a Service (PaaS) [2, 3]. The benefits of using cloud computing are many including pay-per-use, instant on-demand self-service provisioning, speedy elasticity, and resource sharing. Due to the benefits, there is increasing demand for cloud services by enterprises and the scientific applications, which intend calls for the expansion and or building new datacenters. Therefore, the concept of resource allocation (RA) has a meaningful impact on the operations of datacenters. Specifically, in pay-per-use deployments model, which include public, private, community and hybrid Cloud [4].

![Fig. 1. Categorization and Model of Cloud Computing](image)
However, over the years, the high energy consumption of these cloud datacenters have become a major concern as a result of increasing demands of resources and services by enterprise and scientific applications. Due to the large number of equipment contained in datacenters, enamors amount of energy is consumed leading to huge carbon emission [5]. Therefore, the high energy consumption has become a great concern to researchers. For example, Greenpeace (2010) claimed that the Cloud phenomenon may increase the problem of carbon emissions and global warming. The rationale given is that the aggregate demand for computing resources is anticipated to further grow dramatically in the next few years. The technological innovation aimed to reduce overall use of energy is directly related to cost, size and scale of datacenters [6]. Today’s cloud datacenters contain thousands of physical and logical servers for hosting the internet, and other related services that cost millions of Dollars to power them. Even if we don’t consider the financial aspect at present, energy efficiency becomes relevant in design and planning of setting up cloud datacenter [7]. In the datacenter, 75% of energy consumed is because of the linear energy consumed by the PMs [8]. Furthermore, due to the use of high-performance computing with integrated multi-core processors in the PMs, there exists power hunger and dissipation of considerable heat within the datacenter environment [9].

Furthermore, the datacenter energy consumption is proportionate to the resource utilization, beside its virtually considered as the world's largest consumers of electricity [10]. The inefficient usage of the IaaS, poor scheduling policies, and resource under-utilization that are causing the high energy consumption and not their size or low energy-efficiency of the hardware resources [11]. In fact, the utilization level of resources with their corresponding energy used by these datacenters is not trivial [12]. Another reason for this is because less utilized resources waste more energy than those utilized. In this regards, various resource management techniques that are considered to be energy-efficient using classical metaheuristics algorithms have been designed [13-15]. The techniques did not sufficiently prevent underutilization of resources causing the high energy consumption [4]. Similarly, [16-19] have also
proposed scheduling technique using metaheuristics algorithms to scale down the datacenter energy consumption, and resource utilization that includes other related service parameters. Therefore, energy-efficient resource allocation still remains an issue for cloud datacenter service providers. Cloud datacenters offer an abundance of resources, which makes the computing model maintain on-demand resource allocation, such abundance also leads to non-optimal allocation of resources on IaaS which cannot be optimally handled with the existing resource allocation techniques [20, 21]. However, resource optimization is NP-hard problem and all the solution proposed in the literature are based on soft computing method. Since optimality of results for NP-hard problem in soft computing is not provable, thus the focus of the proposed solutions is to optimize the methods to get better results.

Joseph, Chandrasekaran [22], Wu, Tang [23] and Wang, Wang [24] proposed VM placement using Genetic Algorithm (GA) to improve the convergence speed of the GA to produce global optimal solution by the cloud datacenter resource allocation strategy. Furthermore, Particle Swarm Optimization (PSO) algorithm has been explore by various researchers e.g., [25, 26]. Genetic Algorithm for Power-Aware (GAPA) has been developed to resolve the static VM allocation issue in order to improve the datacenter energy-efficiency [13]. Sharma and Reddy [15] proposed a hybrid technique that combined Dynamic Voltage Frequency Scaling (DVFS) with GA to reduce the datacenter energy consumption, increased resource utilization and convergence of the solutions. Alternatively, reducing energy consumption will be realized by turning off or switching PMs that are in the idle state to low-power mode state (i.e., sleep, hibernation) using Order Exchange and Migration (OEM) strategy with Ant Colony System (ACS) [33]. These techniques mapped VMs to PMs randomly and used the fitness evaluation function (objective function) to fit in VMs on PMs with a small number of running application. The jobs are processed based on arrival from 1 to n jobs. These techniques reduce the idle power consumption in the datacenters but has become complicated and difficult to manage due to the imbalance between local and global search of the algorithms which leads to inefficiency in allocating VMs on PMs that also leads
to energy resource wastage. However, these algorithms mostly focus on finding the initial global best solutions and focused only on one-dimensional resource. That is, CPU of PM and the computing requirement of VMs. In choosing a Nature-Inspired technique, there is the need to combine both global and local search methods to balance intensification and diversification. However, a larger solution search space does not always assure a superior optimal solution [21]. This shows the importance of striking a balance between local and global optima, which impacts on the quality of the allocation results. Furthermore, it has been observed in other works, FPA uses Differential Evolution (DE) algorithm to do a local search and also uses static Switching Probability to switch from local to global search space. Experiment results also showed that the local search ability of DE is somewhat limited (Yang and Deb, 2012). This implies that there is need to modify FPA since the central idea of this paper is reducing energy consumption and improving resource utilization in the datacenter. Therefore, the algorithm uses Dynamic Switching Probability (DSP) strategy to find the global optimal solution quickly which increases the convergence speed of the algorithm. Therefore, the adaptation of FPA optimization algorithm to address energy-efficient resource allocation and balancing between global and local search remain as challenging research issue in cloud datacenter environments.

In this context, this paper proposes an Energy-oriented Flower Pollination Algorithm (E-FPA) scheme for virtual machine allocation in cloud datacenter environments. The framework finds the energy oriented optimal solution quickly and balances the exploration and exploitation in the global and local search. The framework can be described majorly in four folds as contribution of the paper:

1) Firstly, an adapted flower pollination model is derived for green computing in cloud datacenter environments.

2) Secondly, a system framework is developed for reducing energy consumption in cloud datacenters focusing on user request model and E-FPA.

3) Thirdly, mathematical analysis of E-FPA is presented based on energy and resource utilization.
Finally, the framework is tested for comparative performance assessments with state-of-the-art techniques considering resource utilization and energy as a metrics within the cloud datacenter settings. The rest of this paper is organized as follows. Section 2 presents the detail of the proposed framework for energy-oriented green computing in cloud datacenter environments using FPA. Simulation setting, and comparative analysis of assessment results are analyzed in section 3. Finally, section 4 present the conclusion followed by future research direction.

2 Energy oriented flower pollination scheme

This section provides the design and development of energy-oriented virtualization scheme using FPA. The scheme addresses the issue of high energy consumption and resource under-utilization due to the imbalance between local and global search which leads to premature convergence and inefficient resource allocation. The scheme contains the following: An overview of FPA, FPA based virtualization in cloud datacenter, and energy-oriented FPA based virtualization.

2.1 Overview of flower pollination modeling for green computing

Flower Pollination Algorithm (FPA) is among the state-of-the-art Nature-Inspired algorithm inspired by the analogy of biological process of pollination [35]. The FPA witnessed significant applications in engineering Ochoa et al. (2014) and various research domain (Abdel-Raouf and Abdel-Baset, 2014; Platt, 2014, Wang and Zhou, 2014). compare to PSO, ACO, CSO, GA, NSGA II and CSA. The performance and effectiveness of FPA are verified using some widely used benchmark problems. The results support its applicability in solving optimization tasks [36]. The summary obtained from FPA evolutionary line, it shows that the algorithm has the readiness, flexibility, capability, and efficiency of being adaptable to solve different types of problems in different NP-hard situations.

Similarly, many researcher have implemented FPA to resolve the NP-hard problems either in continuous or discrete search spaces and found to outperform the compared metaheuristics algorithms [37-39]. The E-FPA deals with the selection of population size \(N\) and a parameter \(P\) which help to select the
value of self-pollination and cross-pollination to take place. The algorithm proceeds by initializing the defined number of population \((N)\), with each one carrying a group of variables which are optimized using the objective function. This algorithm incorporates an indexing term called flower constancy for each population which determines how well their variables minimize the objective function. Based on the flower constancy, the population is queued, and best among them is found as described in algorithm 1.

1) The cross-pollination which is also called biotic pollination is responsible for carrying-pollen to pollinators performing Levy flights movement. The Levy flights is one of the stable distribution which is important in the study of Brownian motion and named after the French mathematician Paul Levy (Borodin and Salminen, 2012). This movement is considered as a global pollination method.

2) The term abiotic and or self-pollination describes the local pollination method.

3) The mean of flower constancy is regarded as probability of reproduction system of the flower that is directly associated to other distinct flowers.

4) Switching probability is employed to control between exploration and exploitation that are commonly known as local and global search. It is defined as \(p [0, 1]\).

The rules above are expressed by three main attributes such as global search, local search, and the switching probability respectively. In FPA, the pollination takes place between two classes except for the fittest function, switching probability, and the levy flight. One of these classes is called the global solution and or global search. In this class, every flower receives single pollen and individual flower drops single pollen gamete only. Consequently, a solution \(x_i\) is equal to a flower and at the same time a pollen. The pollinator’s searches for a solution within the search space to locate the current position of the optimum solution. Therefore, global optimization conforms to the biotic method and or cross-pollination that moves pollen from one location to another different location conforming levy flight law. Levy flight can be express as in Eq. 1.

\[
X_i^f = X_i^f + L (g^* - X_i^f)
\]  
(1)
Here, \( X^*_t \) denotes the \( i^{th} \) solution produced by the pollen at the \( t \) time of iteration. The present solution found within all possible solutions is the best solution and is denoted by \( g^*_t \) during the iteration. \( L \) is also a constraint value indicating movement and scope of the Levy distribution represented as Eq. 2.

\[
L \sim \frac{(\beta+1) \sin \left( \frac{\pi \beta}{2} \right)}{\pi} \times \frac{1}{s^{\beta+1}}, (s \gg s0 > 0)
\]  

Here \( \beta \) is the standard gamma function while the Levy distribution is valid for large steps \( s > 0 \). The second class which is the local search is the composition of flowers after getting the global optimum to find a more optimal solution in the neighborhood. FPA takes local pollinators for searching of a solution within search space due to their effectiveness to locate and obtained a better solution. This class normally finds an improved solution from the current set of solutions evaluated by the objective function. Mathematically it is represented as in Eq. 3.

\[
X^*_t = X^*_t + \varepsilon (X^*_i - X^*_k)
\]

where, \( X^*_i \) and \( X^*_k \) are the pollens of flowers that are alike, but they belong to different species of flowers. \( \varepsilon \) is assumed to be a uniform distribution in \([0,1]\) which become a random walk.

### 2.2 Integration of flower Pollination for resource allocation in cloud computing datacenter

Resource allocation problem is an NP-hard problem. Due to the NP-hard nature of the problems, heuristic algorithms cannot effectively obtain global optimum solutions. Therefore, this research adapted FPA in order to reduce the datacenters challenges in respect to energy and resource utilization. The representation of the FPA in cloud datacenter is shown in Fig 3 which consist of two layers. Each pollen is represented as n-dimensional vector of \( VM \) resources in the \( VM \) layer and each vector is represented as \( p(ij), 1 \leq i, j \leq n \) which consist of random values in the range of \([0,1]\) such that \( 0 < p(ij) < 1 \). The PM layer represent the flower which is a set of PMs where the VMs are placed. The pollen agents serve as the mapping function between \( VMs \) and \( PMs \). The flowers value for mapping \( VM(n) \) on \( PM(m) \) is associated with active \( PM \) within the datacenter. After completion, a unique mapping for each \( VM \) correspond to the \( PM \) for each pollen and flower using the pollen agents that is based on resource optimization.
2.3 Energy oriented virtualization in cloud computing datacenter

The system architecture of resource allocation in cloud environment assigns available resources that are accessible to the various mode. These modes are viewed as a plan for provisioning in datacenter environment using different methods and or schemes as discussed in the introduction section. The methods have not solved the problem of inefficiency of the resource allocation policy that resulted in inadequate resource utilization and energy management in cloud datacenters. However, our method has taken into consideration the problem above and introduces E-FPA a new optimization technique for cloud datacenter to solve the problem. Fig 4 illustrates the components for the proposed energy-efficient resource allocation which is composed of three main entities: Cloud users, service providers, and the datacenter resource management. It shows how the user request is handled by the broker to the Cloud and finally to the datacenter for optimizing VM allocation.

The user’s request is submitted to the Cloud service provider first; then the broker will return the result to the user based on the need, date-line, resource service operation, and capacity/performance management of the available datacenters that they subscribe. When the broker’s request reaches the datacenter, the
cloud information system (CIS) resource manager then looks at the request, compares it with the pool of available resources and then decides. The CIS’s acceptance of any request is based on the system availability. Upon accepting the request, CIS passes it to the scheme for allocation to find global optimal solution. The solution is passed to the resource module which is E-FPA for initial VM placement and monitoring. After which it is placed with the utilized resource that is energy-efficient. In the next stage, the VM manager and scheduler module will identify whether the heterogeneous VMs provide the characteristics of the resource requirement such as reservation, on demand, availability, and allocation. In the following section, we described the user request model, resource mapping and energy models that are applied in realizing energy-efficient resource allocation of cloud datacenter.

![Energy oriented virtualization system components](image)

**Fig 4. Energy oriented virtualization system components**

### 2.3.1 User request model

The user request resources of the datacenter through the broker or cloud provider for their various application needs. The user requests a set of resources known as VMs. Each of the requested resource (VMs) have their required components of performing a task. We denote user request as UR which is the...
subject. The users send one or many requests at a time which are UR \( (A_1, ..., A_n) \) for \( i = 1,2,3, ..., n \) which are executed based on a First-Come-First-Serve (FCFS). \( A_i \) is the components of the VMs. The resource components of VMs include \( \alpha_s^1 \) as CPU, \( \beta_s^1 \) as Memory and \( \gamma_s^1 \) as Storage. The corresponding \( i \) and \( s \) represent the number of resources and their measuring capacity, respectively. Mathematically, we can represent the request as:

\[
A_i \subset UR \text{ and } \alpha_s^1, \beta_s^1, \gamma_s^1 \subset A_i
\]

Therefore, when the user sends a request for only one resource, it will be expressed as in Eq. 4 and 5.

\[
UR^1 = A_i
\]  

(4)

where \( i = 1 \), and \( UR^1 \) is when the resource required is only one.

\[
A_i = (\alpha_s^1 + \beta_s^1 + \gamma_s^1)
\]  

(5)

On the other hand, if the user request is more than one, then the request is expressed as in Eq. 6 and 7.

\[
UR^n = \sum_{i=1}^{n} = A_i = A_1 + A_2 + A_3 + \ldots \ldots A_n
\]  

(6)

\[
= (\alpha_s^1 + \beta_s^1 + \gamma_s^1) + (\alpha_s^2 + \beta_s^2 + \gamma_s^2) + \ldots \ldots ((\alpha_s^n + \beta_s^n + \gamma_s^n))
\]

(7)

2.3.2 Energy and resource utilization model

Given a set of VM = \{vm_i|i = 1,2, \ldots , n\} to be allocated on a set of PMs (Servers) PM = \{PM_j|j = 1,2, \ldots , m\}. Each VM is denoted as a d-dimensional vector of demand resources [40], i.e. \( \text{VM}_i = (A_{i,1}, A_{i,2}, \ldots , A_{i,d}) \). Similarly, each PM is denoted as a d-dimensional vector of capacity resources [41], i.e. \( \text{PM}_j = (B_{j,1}, B_{j,2}, \ldots , B_{j,d}) \). The various PM_j resources considered in this study are 3 including processor (CPU), physical memory (RAM), and storage. Hence, the dimension \( d = 3 \) [40]. Furthermore, the VM allocation has a starting and stopping time, i.e., each VM_i is started at a fixed time.
(S_{t_i}) and execution time(E_{t_i}). Therefore, the overall time expended during the allocation of \( \text{vm}_i \) is mathematically represented as \( S_{t_i} + E_{t_i} \). Where \( A_{i,s} \) is resource capacity \((\alpha_s^1, \beta_s^1, \gamma_s^1)\) requested by the \( \text{VM}_i \) \((1, 2, \ldots, n)\) and \( B_{i,s} \) resource capacity \((\alpha_s^1, \beta_s^1, \gamma_s^1)\) of the \( \text{PM}_j \) \((1, 2, \ldots, m)\). The resource allocation problem has the following (hard) Requirements:

1) Requirements 1: Resource most be compatible with the request

2) Requirements 2: \( \forall \text{VMs} \) request is \( \leq \) the \( \text{PMs} \)'s total resource capacity.

3) Requirements 3: \( \forall \text{VMs} \) each VM is run by a PM at any given time.

4) Requirements 4: Assume that \( a_j(t) \) is the set of VMs that are assigned to a PM.

5) \( \sum_{i=1}^{n} UR^i \) of these assigned VMs is \( \leq \) the \( \text{PMs} \)'s total resource capacity.

\[
\text{For all capacity (} \forall s \text{)} = 1, \ldots, d: \sum_{\text{vm}_i \in a_j(t)} A_{i,s} \leq B_{i,s} \tag{10}
\]

A feasible resource allocation \( R^A \) implies a successful representation of VMs to PMs, i.e., \( \forall i \in \{1,2,\ldots,n\}, \exists j \in \{1,2,\ldots,m\} \), allocated \((\text{VM}_i, \text{PM}_j)\) holds when VM\(_i\) is assigned to physical machine \( \text{PM}_j \). Therefore, the objective function of \( R^A \) is to maximize resource utilization \((RU^{DC})\) and energy efficiency \((DCU_{\text{Energy}})\) of the cloud datacenter. Firstly, maximization of RU of \( \text{PM}_j \) is considered as denoted in Eq. 11.

\[
RU^j_d = \frac{\sum_{i=1}^{n} V_{\text{vm}_i}^d \cdot \text{PM}_j^d}{\text{PM}_j^d} \quad \forall d \in \{\alpha_s^1, \beta_s^1, \gamma_s^1\} \tag{11}
\]

To achieve the total RU of the datacenter, the individual \( \text{PM}_j \) resources are integrated and formulated as thus:

\[
RU^{DC} = \int_{t_1}^{t_2} \frac{\sum_{j=1}^{M_j} V_j^q \cdot \text{PM}_j + \sum_{j=1}^{M_j} V_j^p + \sum_{j=1}^{M_j} V_j^u}{|d| \sum_{j=1}^{M_j} \text{PM}_j} \, dt \tag{12}
\]
We presume that each PM\(_i\) can host any VM\(_i\) and the energy consumption model \(P_j(t)\) of the host PM\(_j\) has a linear relationship with resource utilization (the higher the utilization the higher the energy consumed by the PM) [12]. Lin, Xu [42] use the same model with different resource energy consumption as presented in Eq. 13-15. Table 1 presents energy consumption model of HPG4 and HPG5 servers at various level of utilization. The model for the datacenter PM resource energy consumption is given as follows:

\[
TEC_j^\alpha = \int_{t_1}^{t_2} P(EU(t)) \, dt \quad PM_j \in P
\]  

(13)

where \(EU(t)\) represents the PM utilization at the given time \(t\) and \(P(EU(t))\) represents the power consumption related to \(EU(t)\).

\[
\text{Maximize } \sum_{j=1}^{m} = EU(t)_j
\]  

(14)

\[
TEC_j^\alpha = \sum_{j=1}^{m} = EU(t)_j
\]  

(15)

where \(EU(t)_j\) with \(j = 1,2, \ldots, m\) is the total energy consumption of the PM\(_j\).

\(i \in \{1,2, \ldots, n\}, \quad j \in \{1,2, \ldots, m\}, \quad t \in [0; T].\)

The overall energy consumption for the datacenter can be expressed as presented in Eq. 16.

\[
DC_U^{\text{Energy}} = \int_{t_1}^{t_2} P(EU(t))(\alpha) \, dt + EU(\beta) * P_{\text{max}} + EU(\gamma) * P_{\text{max}}
\]  

(16)

Consequently, the efficiency of \(RA\) which maximizes resource utilization (RU) and energy efficiency of the Cloud datacenter can be mathematically formulated as follows in Eq. 17.
Table 1: Energy consumption by PMs at different load level

<table>
<thead>
<tr>
<th>PM</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP ProG4</td>
<td>86</td>
<td>89.4</td>
<td>92.6</td>
<td>96</td>
<td>99.5</td>
<td>102</td>
<td>102</td>
<td>106</td>
<td>108</td>
<td>114</td>
<td>117</td>
</tr>
<tr>
<td>HP ProG5</td>
<td>93.7</td>
<td>97</td>
<td>101</td>
<td>105</td>
<td>110</td>
<td>116</td>
<td>121</td>
<td>125</td>
<td>129</td>
<td>133</td>
<td>135</td>
</tr>
</tbody>
</table>

2.3.3 Energy oriented flower pollination algorithm

Due to the heterogeneous and large size nature of IaaS cloud and resource management requirement, it is practically impossible to apply FPA directly for resource allocation problem on IaaS cloud due to the large solution space which may take a long time to find an optimal solution. Thus, a new strategy of search operators based on the problem features needs to be re-designed which include the switching probability. Furthermore, it has been observed in other works, FPA uses Differential Evolution (DE) algorithm to do a local search and uses Static Switching Probability to switch from local to global search space. Experiment results also showed that the local search ability of DE is somewhat limited (Yang and Deb, 2012). This implies that there is need to modify FPA since the central idea of the scheme is reducing energy consumption and improving resource utilization in the datacenter. Therefore, the scheme uses Dynamic Switching Probability (DSP) strategy to find the global optimal solution quickly and to increase the convergence speed of the scheme. However, FPA was found to be better compare with other existing resource allocation algorithms and require improvement to meet up with the current increasing number of concurrent users in the cloud datacenter environment. Furthermore, E-FPA requires pollinators at the local search step for efficient search and exploration of solutions within the searching area define by the algorithm. The proposed scheme with the modification is shown in Fig 5. The first step that is performed by the scheme is to find an improved solution from the current solution of the objective function. Any solution that satisfies the constraint of the objective function is considered to be the feasible solution. For example, when using evolutionary algorithm such as GA, candidate solution is de-
signed with a corresponding string of registers popularly identified as a chromosome. Next, every simulation step eliminates the ‘n’ worst scenario solution that have been created and formed a new ‘n’ set of candidate solutions from the best-case scenario of the generated solutions. For every generated solution, there must be a distinctive value that indicates how the solution meets the overall requirement. The goal of resource allocation in the cloud datacenter environment is to allocate the n request to the m available resources to execute the user request with less resource. The details of the algorithm and its implementation are described in the following subsections.

Fig. 5: Operational workflow of the proposed scheme for green computing
3 Implementation of energy oriented flower pollination algorithm

The details of the proposed scheme and its implementation are described in the following section.

3.1 Initialization phase

The first phase of the scheme is the initialization with the aim to find the feasible solution from the current solutions of the objective function, and any solution that satisfies the constraint of the objective function it’s considered to be the best solution. The work uses energy-aware objective function instead of random initialization that reduces the effectiveness of the energy consumption of the optimization. Similar to GA-based optimization method where each solution is designed with a corresponding string of registers popularly known as a chromosome, in this phase each step of a simulation eliminates the $n$ worst scenario solutions that have been created and to form new $n$ set of solutions from the best-case scenario of the generated solutions. For every generated solution, there must be a distinctive value that indicates how the solution meets the overall requirement. If the requirements are not satisfied the algorithm will proceed to the next step. The whole process is iterated until the given stopping criteria are met. The solution is represented by the best flower in the final population. Algorithm 1, shows how the objective function $R^d$ calculates the evaluation value of each pollen defined for resource $i$ and $UR^n$ that aim to reduce the datacenter energy consumption as presented in Eq. 17.

**Algorithm 1 Objective Function**

Require: Total Energy Consumption of Datacenter PMs
Ensure: Allocate VM on PM with efficient energy based on utilization of CPU, Memory, and Storage

1: For each $PM \in$ collection of PMs do
2: $Utilization of PM := PM.getUtilization of PM (CPU; Memory; Storage)$
3: $Power of PM := getPower (PM.getUtilization of CPU)$
4: $Power of PM := getPower (PM.getUtilization of Memory)$
5: $Power of PM := getPower (PM.getUtilization of Storage)$
6: $Energy of Datacenter := Energy of Datacenter + power of PMs (CPU; Memory; Storage)$
7: End for
8: Evaluation Value (Pollen):= 1.0/ power of Datacenter

Springer
3.2 Global search strategy phase

In this phase, it is assumed that each plant contains only one flower, and each flower produces one pollen gamete. Hence a solution $X_i$ is equivalent to a flower and a pollen gamete. Pollinators need to search the whole search space to find the location of the optimum point. Hence, global optimization adapts the biotic and cross-pollinators to play their role more perfectly as they can travel a long distance obeying levy flight rule. Levy flight is more efficient than Brownian in exploring unknown large-scale search space, and this can be express as in Eq. 2.

3.3 Local search strategy phase

The local search phase is the composition of solutions based on Flower Pollination Algorithm after obtaining global optimum solution. The algorithm intensifies the exploitation to find a more optimal solution within the neighborhood structure. FPA needs local pollination for exploitation as they can better exploit the area at which optimum value lies. This phase finds an improved solution from the current solution of the population size as represented in Eq. 3.

3.4 Dynamic switching probability phase

The switching probability $P$ is used to switch between the global pollination and the local pollination in the FPA and the $P$ is always constant. It is assumed that an algorithm should do a more global search at the beginning of the searching process and global search should be less at the end [43]. Therefore, Dynamic Switching Probability (DSP) strategy has been added to adjust the proportion of two kinds of searching process, to balance the local and the global search exploitation and exploration. The enhancement of the switching probability $p$ has been modified according to Eq 18.

\[
P = 0.6 - 0.1 \times \frac{(Max_{\text{iteration}} - t)}{Max_{\text{iteration}}} \tag{18}
\]

where $Max_{\text{iteration}}$ is the maximum iterations of the proposed scheme and $t$ is current iteration time. Special implementation measures of E-FPA including DSP were presented in Algorithm 2 together with the pseudocode.
Algorithm 2 Enhanced Flower Pollination Algorithm

Require: Set of population of \( n \) flowers/pollen gametes with random solutions
Find the best solution \( g^* \) in the initial population

Ensure: Define a switch probability \( P = 0.6 - 0.1 \left( \frac{\text{Max iteration}}{\text{Max iteration}} \right) \)

1: Input: PM list, VM, set of parameters
2: Output: VM allocation
3: Execute: Objective Max \( R^A = (RU^{\text{DC}})(DC_u^{\text{Energy}}) \) // Equation (4.9)
4: Initialize a population of \( n \) flowers/pollen gametes with random solutions
5: Find the best solution \( g^* \) in the initial population
6: Define a switch probability \( P \) using Equation (4.13)
7: While \( (t < \text{Max Generation}) \)
8: For \( i = 1: n \) (all \( n \) flowers in the population)
9: If rand \( < p \),
10: Draw a \((d-\text{dimensional})\) step vector \( L \) which obeys a Levy distribution
11: \[ \text{Global pollination via } X_i^k = X_i^k + L (g^* - X_i^k) \]
12: else
13: Draw \( \in \) from a uniform distribution in [0,1]
14: Randomly choose \( j \) and \( k \) among all the solutions
15: Do local pollination via \( X_{i+1,G} = X_{i,G} + \theta (X_{n-\text{best},G}) + \theta (X_{p,G} - X_{q,G}) \)
where \( \theta = \theta = \epsilon \); 
16: end if
17: Evaluate new solutions
18: Update them in the population
19: Find the current best solution \( g^* \)
20: End while

3.5 Mathematical analysis of E-FPA

The E-FPA is first verified using the benchmarking function proposed by Jamil and Yang [44] and Wang and Zhou [45] which are important to obtain the performance of the optimization algorithm. Table 2 presents the results achieved by the functions used to mathematically compare E-FPA with FPA and ACS using Eq. 19-23.

1) Sphere function
\[ f_1(x) = \sum_{i=1}^n x_i^2 \]  \hspace{1cm} (19)

2) Rosenbrock function
\[ f_2(x) = \sum_{i=1}^n (100 \times (x_i^2 - x_{i+1})^2 + (1 - x_i)^2) \]  \hspace{1cm} (20)
Each of the algorithms were examined in various conditions, i.e., adjusting the value of recursive iteration (100, 200, 1000), keeping the iteration size fixed 40, turning the population size (20, 50, 60), and keeping the number of iteration constant 1000, and using a population size \( n = 25 \) and \( p = 0.8 \) for FPA, crossover probability 0.95 \([35, 45]\), and learning parameters 2 for ACS \([33]\). The result is analyzed based on their performance regarding maximum, average, and standard deviation.

The above algorithm has been run 20 times for each of the above-mentioned settings on a benchmark function, and the conclusive outcome or results are taken from the average of 20 times running of the experiment. This has reduced the impact of the error rate from the experiment. Since in E-FPA the most optimist pollen of the flower can only pass on the information between local and global search using the DSP strategy. As a result, the algorithm converges with high speed against the FPA and ACS. In Fig 6, we plot the convergence of the E-FPA (considering total number of iterations it took to attain the global optimum solution) by changing size of the population from 10 to 50. It has been observed from the experiments, the proposed algorithm converges to global solution fast, retaining a linear link with rising number of population sizes of 10 and 50. Hence, we have selected 50 as size of the population for the conducted experiment. In the proposed scheme, each flower is modelled as \( M \)-dimensional vector. The search space of the flowers is limited to \( I \) and \( I \) is also given in the population size. We observed that the proposed scheme outperforms FPA and ACS. This is because the proposed schemes use DSP strategy that enhances its efficiency. ACS gives the guarantee of convergence, however, the time taken to converge is undefined due to the series of random and casual decisions by the overall scheme while running the experiment. The graph reveals the proposed scheme superiority that solve complex placement problem in cloud computing environment. The average and standard deviation of E-FPA are greater than
FPA and ACS. Therefore, E-FPA presents better results in all the function. The principal feature that ensures the high performance is the introduction of DSP strategy that starts the local search procedure from a feasible solution. Furthermore, this allows the E-FPA find solutions in shorter time.

![Figure 6: Convergence and computational performance](image)

**Table 2: Benchmark Functions Comparison Results**

<table>
<thead>
<tr>
<th>Function</th>
<th>Performance</th>
<th>E-FPA</th>
<th>FPA</th>
<th>ACS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere $f_1(x)$</td>
<td>Maximum</td>
<td>3.97E-04</td>
<td>1.55E-04</td>
<td>9.54E-56</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>2.63E-04</td>
<td>1.70E-04</td>
<td>2.60E-43</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>5.78E+04</td>
<td>1.90E-04</td>
<td>7.87E-51</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>8.61E+04</td>
<td>3.18E-04</td>
<td>8.22E-44</td>
</tr>
<tr>
<td>Rosenbrock $f_2(x)$</td>
<td>Maximum</td>
<td>3.97E-04</td>
<td>1.55E-04</td>
<td>1.32E-03</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>4.04E-04</td>
<td>2.05E-04</td>
<td>4.64E+00</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>2.73E-04</td>
<td>1.24E-04</td>
<td>2.08E+01</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>2.14E+04</td>
<td>1.50E-04</td>
<td>1.71E+00</td>
</tr>
<tr>
<td>Cube $f_6(x)$</td>
<td>Maximum</td>
<td>5.78E+00</td>
<td>2.32E-04</td>
<td>1.03E+00</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>3.18E+00</td>
<td>1.62E-04</td>
<td>5.13E+00</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>2.26E-04</td>
<td>1.18E-04</td>
<td>2.47E+00</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>8.01E+04</td>
<td>2.22E-04</td>
<td>1.32E+00</td>
</tr>
<tr>
<td>Chunk $f_5(x)$</td>
<td>Maximum</td>
<td>5.15E-04</td>
<td>2.52E-04</td>
<td>1.03E-02</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>7.96E-04</td>
<td>3.97E-04</td>
<td>3.77E-02</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>9.71E+04</td>
<td>5.15E-04</td>
<td>1.90E-02</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>8.38E+04</td>
<td>2.43E-04</td>
<td>9.03E-03</td>
</tr>
<tr>
<td>Rastrigin $f_3(x)$</td>
<td>Maximum</td>
<td>5.78E-04</td>
<td>1.90E-00</td>
<td>1.16E+00</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>2.63E-04</td>
<td>1.70E+04</td>
<td>8.76E+00</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>3.97E-04</td>
<td>1.55E+04</td>
<td>3.50E+00</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>4.04E+04</td>
<td>2.05E+04</td>
<td>1.98E-00</td>
</tr>
</tbody>
</table>
4 Evaluation method and result analysis

4.1 Simulation setting

The MulRecCloudsim 3.0.4 is run with the IntelliJ IDEA release version 3.4.0. The schemes are implemented on an Intel CoreTM i7 processor, 2GHz processor speed, 1 terabyte hard disc drive and 8 gigabyte memory. Table 3 shows the parameter settings for the PM and VM used in the experiment. Throughout the simulations time, each VM is assigned a workload randomly trace based on the user request using the same parameter as in [35]. For the sake of simplicity, the PMs are considered to be homogeneous, though heterogeneous configuration can also be simulated. The user request ranges from 1-100 at a time, and the E-FPA is applied at the datacenter on the arrival of a new request from the user. To show the effectiveness of our proposed scheme the $(RU^{DC})$ of (PMs) and $D_C^{Energy}$ of the datacenter are calculated by Eq. 10 and Eq. 17 respectively. The used DSP strategy in the proposed scheme results in global optimum solutions for allocating VMs to PMs which minimizes the datacenter energy consumption. The results of the proposed scheme are compared with Genetic Algorithm for Power-Aware (GAPA) [13] Order Exchange Migration (OEM) ACS [33], and First Fit Decreasing (FFD) [2]. These works have considered resource utilization, number of active PMs and as well their energy consumption. They are clearer and related to the problem we are solving. Furthermore, other researchers have used them in order to compare their work with the same parameter.

<table>
<thead>
<tr>
<th>Cloud Entity</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datacenter</td>
<td>Number</td>
<td>1</td>
</tr>
<tr>
<td>PM</td>
<td>RAM</td>
<td>2048000 MB</td>
</tr>
<tr>
<td></td>
<td>Disk</td>
<td>10000000 MB</td>
</tr>
<tr>
<td></td>
<td>Operating System</td>
<td>Linux</td>
</tr>
<tr>
<td></td>
<td>Bandwidth</td>
<td>1000000000 MB</td>
</tr>
<tr>
<td></td>
<td>Architecture</td>
<td>x86</td>
</tr>
<tr>
<td></td>
<td>VM Manager</td>
<td>Xen</td>
</tr>
<tr>
<td></td>
<td>CPU Power Model</td>
<td>PowerModelSpecPowerX3550XeonX5675</td>
</tr>
<tr>
<td></td>
<td>Storage Power Model</td>
<td>PowerModelStorageSimple</td>
</tr>
<tr>
<td></td>
<td>Memory Power Model</td>
<td>PowerModelMemorySimple</td>
</tr>
<tr>
<td>VM</td>
<td>RAM</td>
<td>2048000 MB</td>
</tr>
<tr>
<td></td>
<td>Bandwidth</td>
<td>0.1GB/s</td>
</tr>
<tr>
<td></td>
<td>MIPS</td>
<td>367 MHz</td>
</tr>
<tr>
<td></td>
<td>Storage</td>
<td>1000000 MB</td>
</tr>
</tbody>
</table>
4.3 Workload type

Experiments were conducted via real data from PlanetLab. The data contains more than a thousand servers with their corresponding components utilizations. The workload consists of 5 days of data as shown in Table 4(a) with several resource demand obtained from the CoMon monitoring project [46]. Datacenter workloads are infrastructure representative in Cloud environment. The data traces of the PlanetLab are accessible and copiously working in CloudSim. Similarly, Amazon EC2 instances has been used in the experiment Table 4(b) shows the four kinds of typical (M3) VM instances suggested by Amazon EC2 considering CPU, memory, and storage as type 1, type 2, and type 3 resources, respectively. For example, request \( (10; 0; 0; 5; 2;) \) represents a user requesting 10 C3.medium VM instances, 0 C3.large VM instance, 0 C3.xlarge VM instance, C3.2xlarge VM instances and 2 C3.4xlarge VM instances.

Table 4. Simulation parameters: (a) planet workload data for 5 days, (b) General purpose (C3) VM instance types offered by Amazon EC2

(a)

<table>
<thead>
<tr>
<th>Data</th>
<th>No. of VMs</th>
<th>Date</th>
<th>Number of PMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workload 1</td>
<td>1085</td>
<td>03/03</td>
<td>800</td>
</tr>
<tr>
<td>Workload 2</td>
<td>896</td>
<td>06/03</td>
<td>800</td>
</tr>
<tr>
<td>Workload 3</td>
<td>1061</td>
<td>09/03</td>
<td>800</td>
</tr>
<tr>
<td>Workload 4</td>
<td>1516</td>
<td>22/03</td>
<td>800</td>
</tr>
<tr>
<td>Workload 5</td>
<td>1078</td>
<td>25/03</td>
<td>800</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Name</th>
<th>CPU</th>
<th>Memory (GB)</th>
<th>Storage (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3.medium</td>
<td>2</td>
<td>3.75</td>
<td>2 x 8</td>
</tr>
<tr>
<td>C3.large</td>
<td>4</td>
<td>7.5</td>
<td>2 x 16</td>
</tr>
<tr>
<td>C3.xlarge</td>
<td>8</td>
<td>15</td>
<td>2 x 40</td>
</tr>
<tr>
<td>C3.2xlarge</td>
<td>16</td>
<td>30</td>
<td>2 x 80</td>
</tr>
<tr>
<td>C3.4xlarge</td>
<td>32</td>
<td>60</td>
<td>2 x 160</td>
</tr>
</tbody>
</table>
4.3 Resource utilization analysis

Fig 5 (a) represents the average resource utilization of CPU, memory, and storages of all the PMs in the datacenter. The total of active PMs inside the datacenter infrastructure is also analyzed. The result shows how the resource utilization is affected as the number of PM increases as depicted in Fig 5(b). The outcomes of results reveal that the E-FPA allocation achieves 95.4% average resource utilization of datacenter as can be seen in Fig 5(c). Whereas in the case of GAPA, OEMACS, and FFD achieve less than 72% average resource utilization of datacenter causing net growth of 23.9% increase in IaaS resources utilization. This improvement in average utilization of PMs is due to the incorporation of DSP strategy that stops the local search from searching specific areas of the search space, thereby making the scheme to explore the neighboring solution. Another reason is the use of DSP strategy that improves the global convergence of E-FPA. The proposed scheme allocates VM on the targeted optimal PM. Overall, the above revealing results justify the benefit of incorporating the DSP strategy in the proposed scheme. This proves that E-FPA is an operational, successful and efficient solution for solving large-scale resource allocation optimization problems.
4.4 Energy consumption analysis

Fig 6(a) shows the E-FPA, GAPA, OEMACS and FFD scheme energy consumption in the Cloud datacenter under different numbers of VM request. The energy consumption of the four schemes increased in different degrees with the increasing number of VM demand by the user. Compared with GAPA, OEMACS, and FFD, E-FPA has the least energy consumption. As can be seen from the results, with an increase of the number of VM request, energy consumption becomes larger and larger. Due to an increasing number of VMs, more PMs will be occupied, bringing greater energy consumption. Also, the performance of E-FPA is always better than the GAPA, OEMACS, and FFD, this is because the proposed scheme can explore the solution space between the local and global search more effectively so
that it can obtain solutions with a reduced number of used PMs compared with other schemes as depicted in Fig 6(b). Therefore, the number of active PMs is reduced and the rate of the energy consumption is remarkably decreases. Hence, total efficiency is increased in the proposed allocation scheme as shown in Table 5. We observed from the table that, the maximum energy consumption is 5015Kwh for E-FPA while the minimum of GAPA is 6520Kwh. OEMACS is 6450Kwh, and FFD is 6756Kwh for the same number of the user request and active PMs. Thus, there is 20.5 % saving in the overall datacenter infrastructure consumption of energy.

Table 5: Average result for the experiment

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Resource Utilization (%)</th>
<th>Active PMs</th>
<th>Energy Consumption (Kwh)</th>
<th>User Request</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-FPA</td>
<td>95.4</td>
<td>450</td>
<td>5015</td>
<td>2000</td>
</tr>
<tr>
<td>GAPA</td>
<td>65.2</td>
<td>736</td>
<td>6520</td>
<td>2000</td>
</tr>
<tr>
<td>OEMACS</td>
<td>71.5</td>
<td>712</td>
<td>6450</td>
<td>2000</td>
</tr>
<tr>
<td>FDD</td>
<td>60</td>
<td>756</td>
<td>6756</td>
<td>2000</td>
</tr>
</tbody>
</table>

Fig. 6. Energy Consumption based on: (a) user request, (b) active PMs
5 Conclusion

This research paper, proposed Energy oriented Flower Pollination Algorithm E-FPA for VM allocation scheme with the goal of reducing datacenter energy consumption and improving on the resource utilization of the physical resources. The models and Algorithm’s pseudo code have been elaborated. E-FPA is more efficient than the GAPA, OEMACS, and FFD schemes in regarding energy consumption and resource utilization. The energy consumption of the datacenter increases when the VM request changes as well. There is an increase in the energy consumption whenever there is an increase in the VM request by users. The allocation scheme uses DSP strategy to find near optimal solution quickly and to balance the intensification and diversification between the global and local search procedure to enhance the efficacy of the allocation scheme. Future research work will be using Multi-Objective approach of FPA to consolidate the datacenter resources.

Acknowledgments

The research is partially supported by the Universiti Teknologi Malaysia research grant global scheduling of IoT applications on cloud reference No: PY/2017/01546 and also University of Gloucestershire research grant School of Computing and Technology, Park Campus. We will also like extend our profound gratitude to the Bauchi State University Gadau Nigeria for their contribution towards the successes of this research.
References


[31] 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


