Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair Resource Allocation in Green Cloud

A. Kalaiselvi[1]* and T. Manoj Prabaharan[2]

[1]*Assistant Professor, Department of Computer Science and Applications, Arul Anandar College (Autonomous), Karumathur, Madurai.*

[2]*Assistant Professor and Head, Department of Computer Science and Applications, Arul Anandar college (Autonomous), Karumathur, Madurai*
**Corresponding Author Email: kalaiselvi@aactni.edu.in*

Abstract

Cloud computing with energy efficient resource allocation is getting acceptance as it is paying significant observation to data management globally. Nevertheless, data centers hosting cloud applications consume vast amounts of energy, resulting in higher amount of operational costs to the environment. As a result, Green Cloud Computing (GCC) solutions are necessitated that can not only reduce operational costs but also minimizes the negative influence of environment factors. Task scheduling in GCC appears to be the most paramount issue to assure cloud connectivity competently and significantly interests the user requirements. Cloud Service Providers (CSPs) employ scheduling and resource allocation techniques that schedule the cloud user requests and manage resource with the view of performing resource allocation by controlling them in full utilization improves the revenue factor. In this work a method called, Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) resource allocation in green cloud computing environment is proposed. First, Antlion and Secant Quasi Newton Multi-objective-based Task Scheduling model, multi-objective learning mechanism (i.e., round trip delay time, CPU utilization and energy consumption) is integrated into Antlion Optimizer to ensure strong global scheduling and avoid falling into local optima. Also better balance between exploration and exploitation is achieved via Secant Quasi-Newton model. Second, Bellman Reinforcement Learning–based Fair Resource Allocation is designed to provide significant resource allocation to the cloud users in the GCC network. Simulation results reveal that the presented model is an effective resource allocation model for providing fair allocation of resources in GCC environments to minimize the energy consumption, average response time and makespan by 34%, 26% and 38% and improves scheduling success by 7% as compared with existing state-of-the-art methods.

Keywords: Green Cloud Computing, Task Scheduling, Antlion Optimizer, Secant Quasi-Newton, Bellman Reinforcement Learning, Fair Resource Allocation

1. Introduction

Cloud computing is an approachable computing model that provides scalability and services on demand to the users using internet. Owing to the elevated acceptance of cloud services, several communities and business enterprises opt for cloud computing environment for deploying scientific workflow appropriate computational resources to achieve the overall execution of tasks while fulfilling the user-defined constraints. Though CC environment has large amounts of advantages like, reducing operational costs, it suffers from high-energy consumption issue. From the above inferences, how to acquire beneficial green computing environment is a significant task is CC environment.

A ε-fuzzy dominance based reliable green workflow scheduling (FDRGS) method was designed in [1]. With this design the reliability of the application and energy consumption were said to be optimized using the ε-fuzzy dominance mechanism. Also with the aid of fast Fourier transform (FFT) and Gaussian Elimination (GE) task graphs it can be inferred that the scheduling algorithm was found to be more effective in optimizing both energy consumption and network lifetime.

Nevertheless, improvements in energy consumption and network lifetime do not guarantee scheduling success. To address on this factor in addition to improving energy consumption, by applying Antlion and Secant Quasi Newton Multi-objective-based Task Scheduling model a significant amount of scheduling success is arrived at.

An Energy Efficient Resource Scheduling using Cultural Emperor Penguin Optimizer (CEPO) method, called EERS-CEPO in Green Cloud Computing (GCC) environment was presented in [2]. The EERS-CEPO in GCC significantly allocated the assigned load between different other resources and therefore circumventing individual resource overload to a greater extent. In addition, the CEPO algorithm was designed by integrating EPO with CA. also optimal resource scheduling in data centers were ensured by means of a fitness function, therefore reducing the operational cost, utilization of energy, generation of heat and cost incurred during the maintenance process.

Despite improvement in energy utilization and operational cost, makespan was not focused. To ensure fair amount of allocation of resources between different cloud user requests, Bellman Reinforcement Learning–based Fair Resource Allocation is designed in this work.

With the elevated upsurge in the utilization of the internet, on-demand for cloud computing provider is also extensively multiplied therefore increasing the on-demand cloud computing services extensively. Cloud service provider construct extensive amount of cloud data center and generates high resources provisioning that in turn necessitates large amount of power to be operated for cloud resources and therefore resulting in large amount of energy consumption also. To reduce total energy consumption for processing, an integrated radio resource allocation and offloading decision optimization model was proposed in [3] taking into consideration the fronthaul and backhaul links, therefore ensuring optimal solution.

Motivated by the application of deep reinforcement learning (DRL) in several domains, a DRL-based strategy that ceaselessly explores states and makes decisions in and adaptive manner, towards long-term cost minimization was designed in [4]. In today’s competitive cloud environment, low energy efficiency will minimize the profit margin of cloud service providers. Owing to this, how to ensure energy efficient VM allocation to attain maximum profit has become a major aspect to be addressed. To focus on this issue, an evolutionary model for VM allocation to reduce energy efficiency of cloud data center was presented in [5].

Cloud resources contribute to huge amount of energy cost and emit carbon to a greater extent. Owing to this reason, cloud services providers necessitate green cloud environment resolution to minimize operational energy cost in addition to reducing environmental influence. To attain green cloud environment an Energy Saving Algorithm was designed in [6]. Yet another method was designed in [7] to address on two aspects, i.e., user association and beam forming issues for ensuring energy efficient allocation of resources.

In [8], Adaptive Four-threshold Energy-aware framework for VM deployment by concentrating on load fluctuation and energy efficiency of VM provisions jointly was proposed. With this mechanism, energy efficiency was said to be improved. The above mentioned works though reduce the energy consumption within data centers to a certain extent, however, they were not found to be suitable in improving scheduling success with minimum makespan for multi-objective functions. Besides, during resource allocation, the average response time for the existing methods is still high and required to be minimized further. The objective of this paper is to perform task scheduling and resource allocation in green cloud computing using Antlion and Secant Quasi Newton Multi-objective-based Task Scheduling model based on round trip delay time, CPU utilization and energy consumption models in green cloud computing environment. Next, Bellman Reinforcement Learning–based Fair Resource Allocation algorithm is applied to the scheduled tasks for ensuring both optimal and fair resource allocation significantly. In this study, the proposed Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) method shows its efficiency to optimize task scheduling and resource allocation compared with other existing resource allocation methods. We have validated our AQNTS-BRF method by performing an extensive performance evaluation study using the CloudSim toolkit. Experimental results show that the AQNTS-BRF method has immense potential as it offers significant energy consumption savings and high potential for the improvement of scheduling success and can minimize the makespan in addition to average response time significantly.

The specific contributions of this paper include the following:

- A literature survey about various existing task scheduling and resource allocation algorithms and an analysis of their advantages and drawbacks are presented.
- An efficient Antlion and Quasi Newton Multi-objective-based Task Scheduling model for task scheduling in green cloud computing environments is proposed.
- An algorithm for resource allocation in green cloud computing environments inspired by reinforcement learning called, Bellman Reinforcement Learning–based Fair Resource Allocation is proposed.
- Performance analysis of the proposed method and an evaluation of the algorithm with respect to other existing algorithms for scheduled resource allocation are presented.

2. Related works

Cloud computing refers to the collection of computer system resources that are provision in an arbitrary fashion to ensure services to concerned users on the basis of demand access. With the increase in the requirements of customers' services offered by data centers, energy consumed by data centers are also said to be increased significantly.

A trade-off between energy being consumed and system performance in multi-clouds system was presented in [9]. Taking into consideration the cost incurred in transmission and corresponding cost involved in execution process, a dynamic optimization problem with the purpose of reducing power consumption was formulated. Also, a Collaborative Task Offloading and Resources Allocation Algorithm (CTORAA) were designed to determine task offloading and energy harvesting. Despite effectiveness observed in energy harvesting, migration rate were also found to be high. To address on this issue, a resource management algorithm for virtualized data centers [10] that not only optimizes the servers to ensure dynamic workloads but also at the cost of minimum migration was ensured.

A survey of different techniques that investigated the energy efficiency and problem-solving techniques utilized for minimizing power consumption in data centers was investigated in [11]. The quality of service in green cloud computing is said to be compromised due to VM placement. It necessitates in designing a mechanism that enhances energy by proportionate resources in a smart way. Hence, in [12], a detailed review and analysis based on certain performance parameters and metrics for VM migration like, optimization of energy and consolidation was made.

Most of the prevailing IT-based businesses use cloud computing technology. One of the growing technologies is cloud computing that persistently add numerous services to sustain their likelihoods of competition and meet the increasing necessitates of customers. With respect to the above observations, how to design a green computing is still considered as a major challenge in cloud computing environments

An energy-efficient resource allocation and optimal migration policy was designed in [13]. Yet another energy-efficient resource allocation scheme based on floating point calculations per second employing joint channel matrix sparseness and normalized water-filling resource allocation algorithm was proposed in [14], therefore improving energy efficiency to a greater extent. Nevertheless, the processing time involved in the overall process was not concentrated.

In [15], a nested Particle Swarm Optimization (TSPSO) via multi-objective function to optimize both energy and processing time was proposed. With this an optimal balance was said to be ensured. An energy-efficient hybrid (EEH) framework for enhancing the significant of electrical energy consumption in data centers was proposed in [16]. Here, the power consumption involved in scheduling decisions employing migration algorithm was designed that in turn not only reduced average execution time but also improved throughput. Yet another green technology using Long Term Evolution (LTE) network was presented in [17].

A detailed performance comparison of several tasks scheduling in green cloud computing environment based on heuristic algorithm was designed in [18]. Reducing energy consumption has become a critical issue in today data centers. Minimizing the frequency and volume of required physical and VM results in energy-efficiency mechanisms. In [19] to discard the limitations of VM migration, a static VM placement technique was proposed that places VMs on hosts in a Worst-Fit-Decreasing (WFD) fashion was presented. Neural network and policy gradient was applied in [20] to address on the volatility aspect of green cloud data computing.

It is evident from the above discussion that several researchers have designed the task scheduling and resource allocation problem in the past few years. However, most of the existing methods do not consider multi-objective as constraints while performing scheduling and allocating the resources in green cloud computing environment. To address on this angle a novel method called, Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) resource allocation in green cloud computing environment is proposed in this work. The elaborate description of AQNTS-BRF method is provided in the following sections.

3. Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) resource allocation in green cloud computing environment

This section presents the multi objective-based task scheduling and resource allocation (MUTS-RA) method called, Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) resource allocation in green cloud computing environment which is designed for data execution and related services where real-time responses are required. Figure 1 shows the block diagram of AQNTS-BRF method.

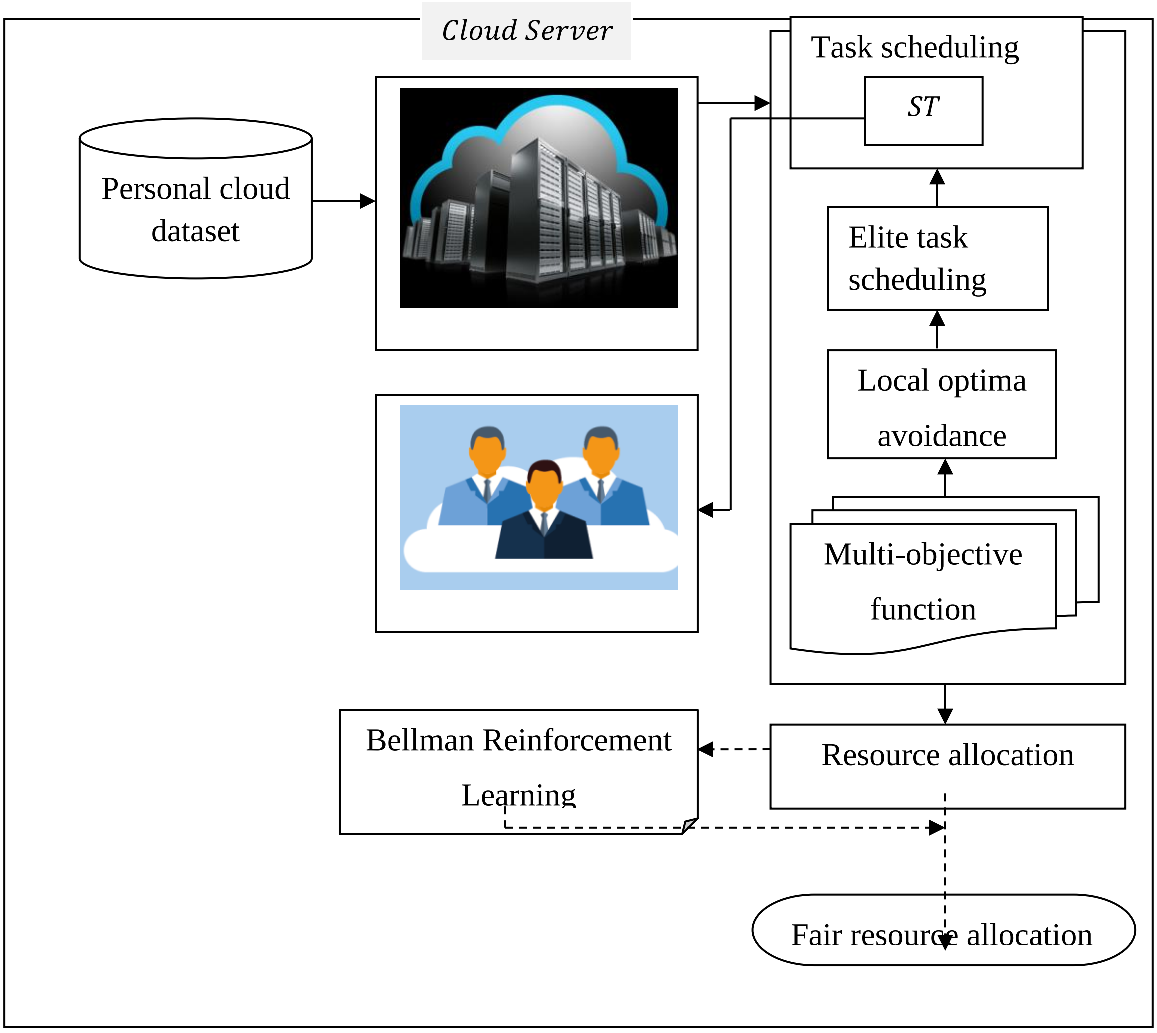


Figure 1 Block diagram of AQNTS-BRF method

As illustrated in the above figure, the proposed AQNTS-BRF method first schedules the cloud user requested tasks using Antlion and Quasi Newton Multi-objective-based Task Scheduling model. Here scheduling is done by considering the multi-objective formulation (i.e., round trip delay time, CPU utilization and energy consumption) so that the overall task energy consumption can be minimized and scheduling success can be improved as much as possible. Further minimizing the makespan and average response time is considered as the most critical characteristic as far as fair resource allocation is concerned. Following which Bellman Reinforcement Learning–based Fair Resource Allocation is designed to ensure fair allocation of resources to the scheduled tasks.

3.1 Green Cloud Computing Network model

Each cloud consists of one or more data centers with each data center possessing several VMs that addresses the requirements of cloud users' requests. From this angle, each data center possesses a set of physical machines that denotes the platforms on which VMs are located and run. Let us further assume that '$S = \{S_1, S_2, \ldots, S_s\}$' represents the servers and '$VM = \{VM_1, VM_2 \ldots, VM_{vm}\}$' denotes '$vm$' virtual machines positioned on the server 's' respectively. Finally, each 'VM' also posses both computing and storage resources '$R = \{R_1, R_2, \ldots R_r\}$', to name a few are CPU, memory and so on that have to be allocated to the virtual machine 'VM'. During task scheduling process, the cloud user requested tasks are allocated with the available data centers '$DC = \{DC_1, DC_2, .., DC_d\}$'. Here, each data center 'DC' is associated with 'm' number of Processing Elements (PEs) to execute cloud user requested tasks and each PEs are associated with a triplet '$< RTDT, P(CPU_{util}), EC >$' round trip delay time, CPU utilization and energy consumption of each Processing Elements respectively. The proposed method comprises of 'm' independent tasks '$T = \{T_1, T_2, \ldots, T_m\}$' acquired in a random manner from 'n' different cloud users '$U = \{U_1, U_2, \ldots, U_n\}$'. The corresponding task-cloud user matrix is structured as given below.

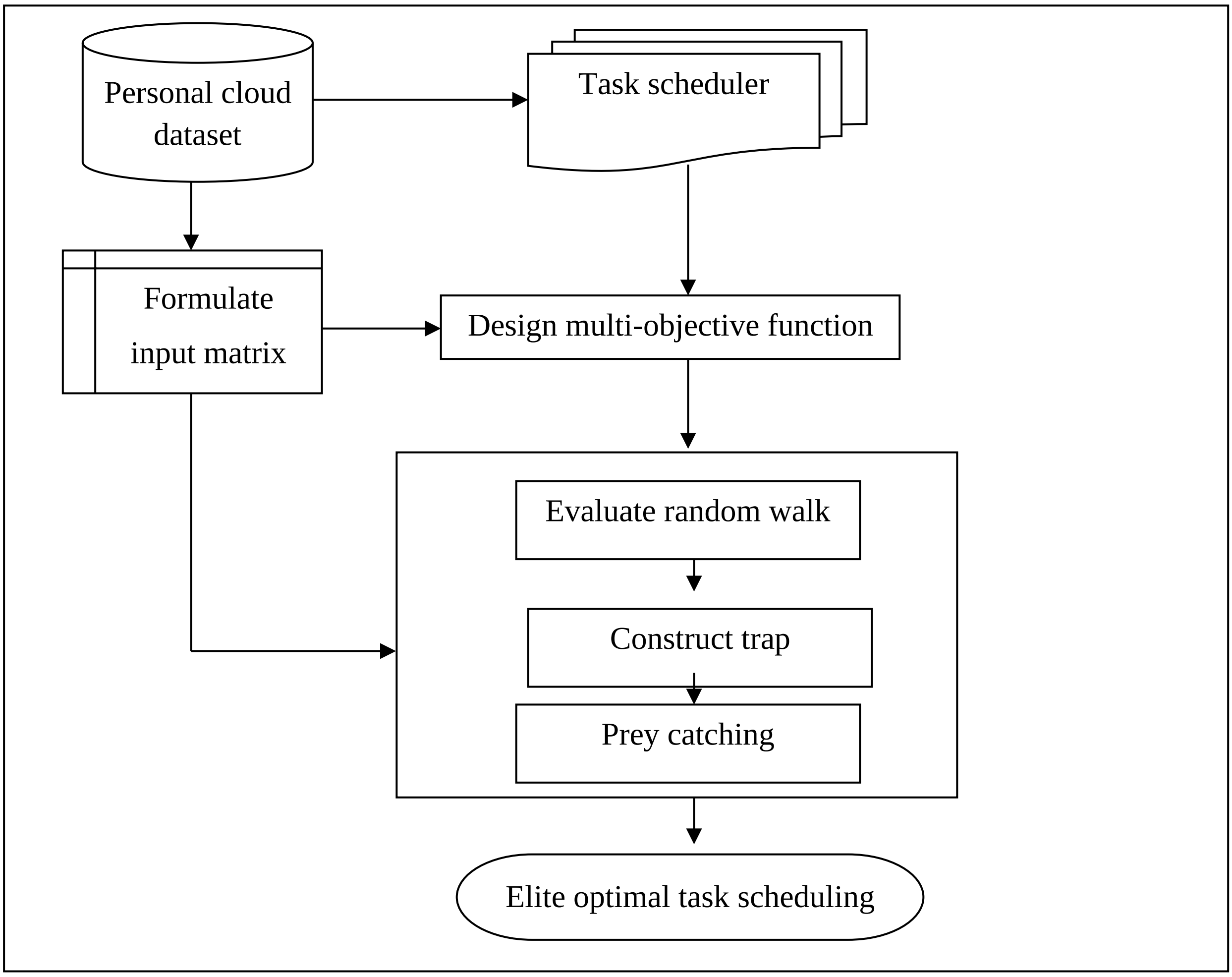
$$X(t) = \begin{bmatrix} U_1T_1 & U_1T_2 & \ldots & U_1T_m \\ U_2T_1 & U_2T_2 & \ldots & U_2T_m \\ \ldots & \ldots & \ldots & \ldots \\ U_nT_1 & U_nT_2 & \ldots & U_nT_m \end{bmatrix} \quad (1)$$

With the above cloud user requested task matrix '$X(t)$' is formulated, by utilizing multi-objective under the constraint of round trip delay time, CPU utilization and energy consumption the optimal tasks scheduling and resource allocation are modeled.

3.2 Antlion and Quasi Newton Multi-objective-based Task Scheduling

The main objective of the green cloud computing is to improve the performance efficiency of system resources by minimizing the amount of total electrical power consumed by the concerned resources. Several resources in cloud data center to name a few being, hardware, data storage units, power management units are laborious and cumbersome to control as far as power consumption is concerned owing to the reason that the backend management mechanism do not take into consideration the energy and environmental influence of power allocation. In addition, the computational resources have to be used in an efficient manner to minimize energy waste. As a result, these three factors, energy efficiency, CPU utilization and round trip delay time are required to be optimized and robust as far as green data center solutions are concerned.

Initially, to enhance the global scheduling ability, multi-objective learning mechanism is proposed through the analysis of round trip delay time, CPU utilization and energy efficiency. Then, the multi-objective learning mechanism is effectively integrated into Antlion to obtain strong global scheduling and avoid falling into local optima. Finally, to keep a better balance between exploration and exploitation, a Secant Quasi-Newton model with strong exploitation potentiality is incorporated into Multi-objective Learning-Antlion. Figure 2 shows the structure of Antlion and Quasi Newton Multi-objective-based Task Scheduling model



As illustrated in the above figure, with the records in the personal cloud dataset provided as input, the task scheduler performs the task of elite optimal task scheduling. This is performed by first formulating input matrix, following which multi-objective function is designed for both scheduling cloud user requested tasks and allocating resources in green cloud computing. Finally, optimal task scheduling is performed via Antlion and Quasi Newton function. The first objective remains in reducing the round trip delay time 'RTD_{ij}' of task 'T_i' in 'VM_j' assignment. This is mathematically formulated as given below.

$$RTDT_{ij} = \left[\frac{Size(T_i)}{BW}\right] + D + \left[\frac{Ins(T_i)}{Ins\,(VM_j)}\right] + D \tag{2}$$

From the above equation (2), the round trip delay time '$RTDT_{ij}$' is evaluated based on the task size '$Size(T_i)$', bandwidth 'BW', delay 'D', number of tasks instructions '$Ins(T_i)$' to be executed, number of instructions executed by virtual machine '$Ins\left(VM_j\right)$' respectively. The second objective remains in minimizing the total power consumption by switching passive servers to sleep mode and vice versa. The power model is defined as given below.

$$P(CPU_{util}) = P_{PS} * P_{max} + (1 - P_{IS}) * P_{max} * CPU_{util} \tag{3}$$

From the above equation (3), the power consumption with respect to CPU utilization '$P(CPU_{util})$', is measured based on the maximum power consumption when the CPU is completely used 'P_{max}', power consumption of the passive server 'P_{PS}' and utilization of CPU 'CPU_{util}' respectively. CPU utilization may change with respect to owing to the variation in the workload at the data center. As a result, the CPU utilization in our work is modeled as a function of time denoted as '$CPU_{util}[t]$'. Third, the energy consumption is formulated as given below.

$$EC = \int_{low}^{up} P(CPU_{util}[t])dt \tag{4}$$

From the above equation (4), the energy consumption 'EC' is modeled based on the CPU utilization with respect to time '$CPU_{util}[t]$' ranging between lower 'low' and upper 'up' bound respectively. Finally, multi-objective function is formulated as given below.

$$Obj = Min\left(\sum_{i=1}^{m}\sum_{j=1}^{n} RTDT_{ij}\right) + Min\,[P(CPU_{util})] + Min\,(EC) \tag{5}$$

$$X_{obj}(t) = \begin{bmatrix} Obj(U_1T_1, U_1T_2, \dots, U_1T_m) \\ Obj(U_2T_1, U_2T_2, \dots, U_2T_m) \\ \dots \\ Obj(U_nT_1, U_nT_2, \dots, U_nT_m) \end{bmatrix} \tag{6}$$

From the above equations (5) and (6), the multi-objective function 'Obj ' is formulated by minimizing the round trip delay time '$Min\left(\sum_{i=1}^{m}\sum_{j=1}^{n} RTDT_{ij}\right)$', power consumption of CPU utilization '$Min\,[P(CPU_{util})] +$' and the energy consumption '$Min\,(EC)$' respectively. With the above multi-objective functions, the main random walk of ants (i.e., the random cloud user requested tasks placed in the cloud server) is obtained from the equation (1). Following which the random walk of ants (i.e., random cloud user requests) is formulated as given below.

$$X_i(t) = \frac{(X_i(t) - Min_i) * \left(Max_i - Min_i(t)\right)}{Max_i(t) - Min_i} \tag{7}$$

From the above equation (7), the random cloud user requests placed in the server is modeled based on the minimum of random walk 'Min_i' of '$i - th$' request (i.e., minimum of requests placed by cloud user), maximum of random walk ‘Max_i’ of ‘$i-th$’ request (i.e., maximum of requests placed), minimum of ‘$i-th$’ request at ‘$t-th$’ iteration ‘$Min_i(t)$’ (i.e., minimum of requests placed by cloud user at a specific time instance) and maximum of ‘$i-th$’ request at ‘$t-th$’’ iteration ‘$Max_i(t)$’ (i.e., maximum of requests placed by cloud user at a specific time instance) respectively. The next step remains in the design on constructing trap. To be more specific no cloud user requests has to be performed by the cloud server as its requirement are optimal, i.e., avoiding getting trapped into local optimality. This function is mathematically represented as given below.

$$Min_j(t) = X_j(t) + Min(t) \quad (8)$$

$$Max_j(t) = X_j(t) + Max(t) \quad (9)$$

From the above equations (8) and (9), to avoid getting trapped ‘$Min_j(t)$’ and ‘$Max_j(t)$’ are formulated based on the minimal of all the cloud user requests ‘$Min(t)$’ and the maximal of all the cloud user requests ‘$Max(t)$’ respectively. Following which task scheduling between requests is performed as formulated below.

$$X_j(t) = X_i(t), iff \left(X_i(t) > f\left(X_j(t)\right)\right)$$

(10)

From the above equation (10), optimal task scheduling occurs when the ants (i.e., cloud user request) gets fitter than its associated antlions (i.e., associated cloud user request). Finally, to ensure balance between the exploitation and exploration, Elite-based Secant Quasi-Newton function as given below.

$$X_{j+1} = X_j - f'(X_j) * \frac{X_j - X_{j-1}}{f'(X_j) - f'(X_{j-1})}$$

(11)

As given in the above equation (11), ant’s movement are said to be influenced by either probability or the best antlion or elite (i.e., optimal ensuring three functions). The range of all random walks reduces proportionally to updated iterations. In this manner, multi-objective functionality with global scheduling by balancing between exploitation and exploration is achieved in an efficient manner. The pseudo code representation of Antlion and Quasi Newton Multi-objective-based Task Scheduling is given below.

```
Input: Dataset ‘$DS$’, User ‘$U = \{U_1, U_2, \dots, U_n\}$’, Task ‘$T = \{T_1, T_2, \dots, T_m\}$’, Data Center
‘$DC = \{DC_1, DC_2, \dots, DC_d\}$’
```

```
Output: Min-max Task scheduling (i.e., minimum energy consumption and maximum scheduling success)
Step 1: Initialize'$m$', '$n$', '$d$', host '$k$', Task Scheduler '$TS$', Virtual Machine '$VM = \{VM_1, VM_2 \ldots, VM_{vm}\}$', Server '$S = \{S_1, S_2, \ldots, S_s\}$', '$s$', resources '$R = \{R_1, R_2, \ldots R_r\}$', '$r$'
Step 2: Initialize '$P_{max} = 250W$', bandwidth '$BW$', delay '$D$', lower bound '$low = 0.5$', upper bound '$up = 1$', '$t = [0.5,1]$'
Step 3: Begin
Step 4: For each Dataset '$DS$' with User '$U$', Task '$T$', Data Center '$DC$'
//initialize random walk
Step 5: Formulate random walk of ants as given in equation (1)
//multi-objective function
Step 6: Formulate round trip delay time model as given in equation (2)
Step 7: Formulate CPU utilization as given in equation (3)
Step 8: Formulate energy consumption as given in equation (4)
Step 9: Formulate multi-objective function as given in equation (5) and store it in matrix as given in (6)
Step 10: Evaluate random walk as given in equation (7)
//trap construction
Step 11: Formulate avoiding of getting trapped into local optimality as given in equations (8) and (9)
//catching prey
Step 12: Perform task scheduling between requests as given in equation (10)
//Elite-based Secant Quasi-Newton function
Step 13: Formulate optimized task scheduling as given in equation (11)
Step 14: Return scheduled tasks '$ST$'
Step 15: End for
Step 16: End
```

Algorithm 1 Antlion and Secant Quasi Newton Multi-objective-based Task Scheduling

As given in the above algorithm, the objective remains in improving the scheduling success with minimum energy, a multi-objective function based Antlion and Quasi Newton task scheduling is modeled. Three processes are designed. First, random walk of ants or cloud user requested tasks placed in cloud server is formulated. Next, multi-objective functions are modeled and finally, the multi-objective functions are stored for the corresponding random walk of ants or cloud user requested tasks. Second, Antlion-based task scheduling is formulated where random walk is initialized, next, trap construction is modeled to ensure global scheduling and avoid failing into local optimization. Finally, balance between exploitation and exploration is ensured so that equal chances are ensured for all the cloud user request tasks being scheduled in an optimal manner via Elite-based Secant Quasi-Newton function. In this manner, task scheduling is achieved with minimum energy consumption and maximum scheduling success.

3.3 Bellman Reinforcement Learning –based Fair Resource Allocation

Loads to data centers hosted by cloud platforms are increasing over the past few years. Therefore, it becomes essential to design an allocation model that provides competently better quality of services to ensure smooth balance between requirements and availabilities. In this work multi-objective task scheduling-based resource allocation between cloud backend data centers is presented. The model takes into account the multi-objectives into consideration along with the utilization of resources that satisfy all objectives towards green cloud computing.

The novelty of the multi-objective task scheduling-based resource allocation lies in the application of Bellman Reinforcement Learning-based Fair Resource Allocation for dynamically allocating the resources. This model embeds Bellman's principle of optimality for the consideration of tradeoff between similar amounts of resources required by cloud users' requests. Bellman Reinforcement Learning–based Fair Resource Allocation in our work learns the process (i.e., dynamic and optimal resource allocation) by interactions with dynamic environment (i.e., green cloud computing) that in turn generates optimal control policy for given set of states (i.e., set of scheduled tasks) without requiring domain knowledge of the environment. Figure 3 shows the block diagram of Bellman Reinforcement Learning-based Fair Resource Allocation model.

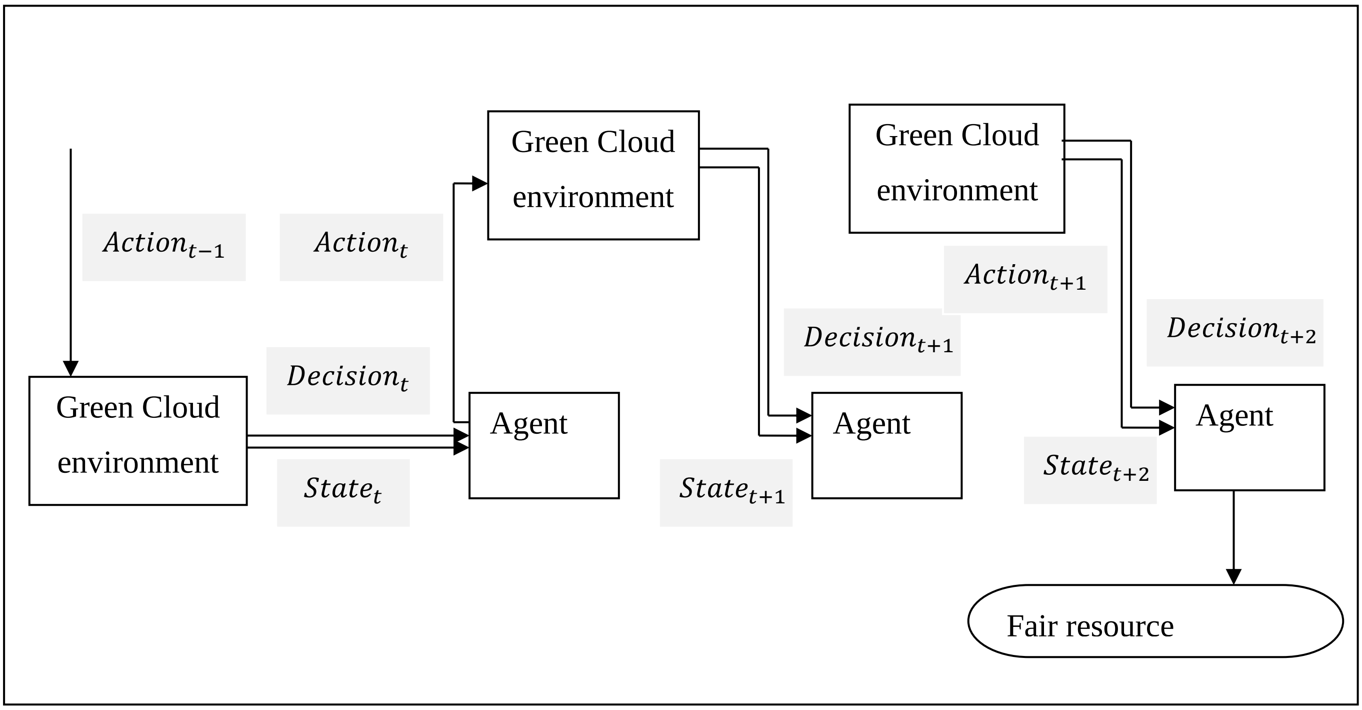


Figure 3 Structure of Bellman Reinforcement Learning-based Fair Resource Allocation model

As illustrate in the above environment, with the scheduled tasks obtained as input, the objective remains in allocating the resources in a fair manner. Let '$R = \{R_1, R_2, \dots, R_r\}$' represent the set of '$r$' available resources that should process 'st' independent scheduled tasks '$ST = \{ST_1, ST_2, \dots, ST_{st}\}$'. The main objective remains in enhancing the multi-objectives (as provided in algorithm 1) of the data center and minimize makespan so as to attain dynamic and optimal resource allocation. This is formulated in our work by means of state space, action space and Bellman optimal reward respectively.

In green cloud computing data center environment, 'N' hosts are employed to serve as resources for cloud user requests and each host is modeled in the form of state 'S' with the state set of the green cloud computing environment is represented as '$SS = \{SS_1, SS_2, \dots, SS_N\}$'. Each state set '$SS_i$' for time '$t$' has four dimensions, energy consumption of data center '$DCEC_t$', number of hosts in data center 'DCH_t', CPU utilization of host '$CPUH_t$', and demand of virtual machine 'VM_t' respectively. Following which the probable action set '$AS = \{AS_1, AS_2, \dots, AS_M\}$' for any host is said to be either 'ASSIGN' or 'NOT ASSIGN' based on the data center energy and current host utilization.

The dynamic and optimal resource allocation model is based on utilizing a reward function that evaluates the multi-objective parameters specified in the task scheduling model as well as the four dimensions utilization function of every cloud user request tasks. Bellman's principle of optimality is employed for mapping of these parameters into reward value. The reward value function is represented in the form of a quintuple '$(SS, AS, TPM, Rew, \delta)$', where '$SS$' represents the state set, 'AS' denotes the action set, with the transitional probability matrix represented as 'TPM', Bellman Optimal Reward function represented as 'Rew' and finally a discount factor 'δ' respectively. The transitional probability matrix and reward matrices are formulated as given below.

$$TPM_{s,s'}^{q} = TPM_r(SS_{t+1} = s' | SS_t = s, AS_t = a)$$

(12)

$$Rew_s^a = TPM_r(Rew_{t+1} = r | SS_t = s, AS_t = a)$$

(13)

From the above equations (12) and (13), the reward value ‘Rew_t’ is evaluated for the corresponding action ‘AS_t’ upon occurrence of the transition probability ‘$s \rightarrow s'$’, where ‘$s \in SS_t$ ’, ‘$a \in AS_t$’, and ‘$r \in Rew_t$’ respectively. However, in case of similar numbers of resources are requested via cloud user requests, then fairness is said to be compromised. The resource allocation in Energy Efficient Resource Scheduling using Cultural Emperor Penguin Optimizer (CEPO) algorithm [2] was characterized by a trade-off between efficiency and fairness. To address on this aspect, Bellman Optimal Reward function is employed with which maximization of efficiency subject to constraints is designed in such a manner so as to ensure that the decision is acceptably fair.

The Bellman Optimal Reward (set of decisions) has the characteristic that whatever the initial state (i.e., cloud user requests, scheduled tasks) and decisions are (i.e., allocation of resources for the scheduled tasks), the remaining decisions must account for optimality with regard to state set resulting from the first decision. The Bellman Fair Reward function is formulated as given below.

$$r(s,a,s') = \sum r * TPM_r(r|s,a,s') = \frac{\sum r * TPM_r(s',r|s,a)}{TPM_r(s'|s,a)}$$

(14)

$$f_{opt}(Dec_i) = \max\left[Ret\left(Dec_n\left(r(s,Dec_i,s')\right)\right)\right] + f_{opt} - Tf(Dec_i,Dec_n)$$

(15)

From the above equation (15), the Bellman Fair Reward function for the initial state ‘$f_{opt}(s_i)$’ is modeled based on the return due to decision ‘$Ret\left(Dec_n\left(r(s,Dec_i,s')\right)\right)$’, transfer function providing the resulting state ‘$Tf(Dec_i,Dec_n)$’ with initial set of decisions ‘$\{Dec_i\}$’ respectively. The pseudo code representation of Bellman Reinforcement Learning–based Fair Resource Allocation is given below.

```
Input: Dataset ‘$DS$’, User ‘$U = \{U_1, U_2, \ldots, U_n\}$’, Scheduled Task ‘$ST = \{ST_1, ST_2, \ldots, ST_{st}\}$’, Data Center ‘$DC = \{DC_1, DC_2, \ldots, DC_d\}$’, Resources ‘$R = \{R_1, R_2, \ldots, R_r\}$’
Output: optimal response time and makespan improved resource allocation
Step 1: Initialize ‘$st$’, ‘$r$’, ‘$N$’, ‘$M$’, discount factor ‘$\delta = (0,1)$’
Step 2: Begin
Step 3: For each Dataset ‘$DS$’ with User ‘$U$’, Scheduled Task ‘$ST$’, Data Center ‘$DC$’
//State space
```

```
Step 4: Formulate energy consumption of data center ‘$DCEC_t$’, number of hosts in data center ‘$DCH_t$’, CPU utilization of host ‘$CPUH_t$’, and demand of virtual machine ‘$VM_t$’
//Action space
Step 5: Formulate action set ‘$AS = \{AS_1, AS_2, \dots, AS_M\}$’
// Reward function
Step 6: Formulate transitional probability matrix as given in equation (12)
Step 7: Formulate reward as given in equation (13)
// Bellman Optimality
Step 8: Formulate Bellman Fair Reward function as given in equations (14) and (15)
Step 9: Return fair resource allocated
Step 10: End for
Step 11: End
```

Algorithm 2 Bellman Reinforcement Learning–based Fair Resource Allocation

As given in the above algorithm, with the objective of improving makespan and average response time involved in resource allocation, reinforcement learning is first designed with the scheduled tasks acquired as input. Initially, the state space considering the energy consumption of data center, hosts in data center, host CPU utilization, and demand of virtual machine are first formulated. Second, the action set are designed, following which, a transitional probability matrix and reward functions are modeled. Finally, to ensure fair allocation of resources between scheduled cloud user requests, Bellman Fair Reward function is applied that in turn not only ensures optimality but also minimizes the makespan and average response time significantly.

4. Experimental settings

This section is all about computational experiments that are utilized in evaluating the performance of the Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) method. The AQNTS-BRF method has been simulated by means of CloudSim tool with the base platform performed on JAVA. All the given below set of experiments are validated on a PC with Intel(R) Core(TM) i5-457, 4 CPU @ 2.9 GHz, a RAM of 8GB and 64-bit Windows OS as its configuration. The results are compared with existing methods [1] and [2]. The parameters employed are given in the form of table 1. In this experiment Personal Cloud Dataset obtained from http://cloudspaces.eu/results/datasets is used in order to check the performance of the AQNTS-BRF method. Also, detailed analysis by simulation the AQNTS-BRF method in green cloud computing environment with existing methods, -fuzzy dominance based reliable green workflow scheduling (FDRGS) [1] and Energy Efficient Resource Scheduling using Cultural Emperor Penguin Optimizer (CEPO) algorithm, called (EERS-CEPO) [2] are briefed.

Table 1 Parameters and description

S. No	Parameters	Description
1	Number of data center	5
2	Number of host	10
3	Host memory	10 GB
4	Host bandwidth	2800Mbps
5	Number of virtual machines	50
6	Bandwidth	250 – 1500
7	Number of tasks	5000 – 50000

4.1 Performance metrics

In order to compare the efficiency of the Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) method with a detail comparative analysis of ε-fuzzy dominance based reliable green workflow scheduling (FDRGS) [1] and Energy Efficient Resource Scheduling using Cultural Emperor Penguin Optimizer (CEPO) algorithm, called (EERS-CEPO) [2] algorithms we use several metrics to evaluate their performance. The first metric is the total energy consumption by the physical resources of a data center caused by the application workloads using various methods. Energy Consumption refers to the energy consumed for the corresponding cloud user requested tasks with by all the hosts in the virtualized Cloud. The energy consumption of each host is computed using

$$EC = \sum_{i=1}^{m} EC_{Trans}(T_i) + EC_{Exec}(T_i)$$

(16)

From the above equation (16), energy consumption 'EC' is measured based on the energy consumed during the transmission for a specific task '$EC_{Trans}(T_i)$' and the energy consumed during the process of execution for a specific task '$EC_{Exec}(T_i)$' respectively. It is measured in terms of joules (J). The second performance metric is called the scheduling success which is defined as the percentage ratio of tasks submitted to the scheduled task. The

Scheduling Success 'SS' of a task 'T_i' scheduling success is mathematically defined as given below.

$$SS = \sum_{i=1}^{m} \frac{T_{SD}}{T_i} * 100$$

(17)

From the above equation (17), the scheduling success 'SS' is formulated based on the total number of tasks submitted to the cloud server 'T_i' and number of tasks scheduled according to the deadlines 'T_{SD}' respectively. It is measured in terms of percentage (%). The third metric is the makespan. Makespan refers to the maximum completion time of cloud user requested task. The makespan is mathematically stated as given below.

$$MS = Max[CT \sum_{i=1}^{m} T_i]$$

(18)

From the above equation (18), makespan 'MS' is measured on the basis of the maximum 'Max' completion time 'CT' with respect to corresponding cloud user requested task 'T_i' respectively. The last performance metric is the average response time. The average response time refers to the time consumed in responding cloud user request whenever there arises' a request for resource. To be more specific, the average response time is measured as the ratio of total service time to waiting time of entire task. It is mathematically formulated as given below.

$$ART = \frac{\sum_{i=1}^{m} WT(T_i)}{TST}$$

(19)

From the above equation (19), the average response time 'ART' is measured based on the total service time 'TST' and the overall waiting time of all the tasks in queue '$WT(T_i)$' respectively. It is measured in terms of milliseconds (ms).

4.2 Case scenario 1: Energy consumption

Energy efficiency is considered as one of the key aspects as far as resource allocation in green cloud computing environment is concerned. When the lower utilization threshold increases, the energy consumption of the system is also rapidly reduced. The statistical analysis of the energy consumption comparison of the three methods, AQNTS-BRF, FDRGS [1] and EERS-CEPO [2] is given in table 2. It can be observed that there is a significant difference between the three methods and our proposed AQNTS-BRF method consumes the least energy in most cases.

Table 2 Tabulation for energy consumption

Number of tasks	Energy consumption (J)		
	AQNTS-BRF	FDRGS	EERS-CEPO
5000	2500	4500	5500
10000	2810	4800	5800
15000	3100	5100	6420
20000	3410	5210	7320
25000	4100	5310	7910
30000	4570	5550	8000
35000	5210	6120	8110
40000	5830	6900	8350
45000	6210	7210	8520
50000	6810	7830	9320

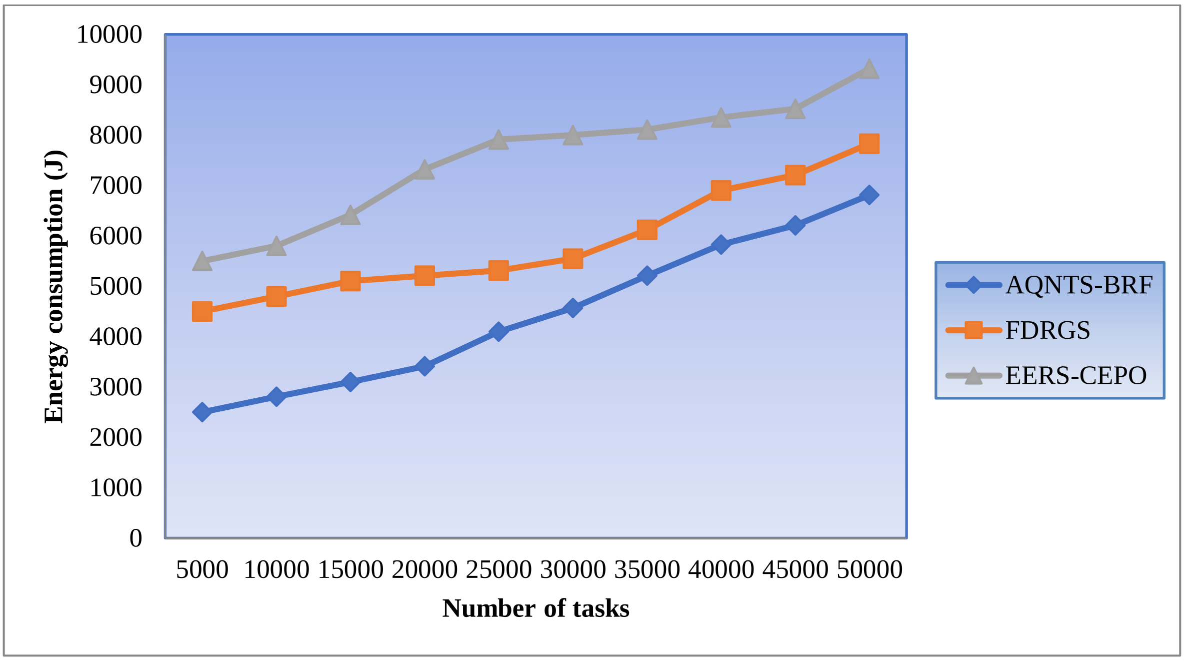


Figure 4 Comparison of the energy consumption of the four scheduled resource allocation methods

Figure 4 depicts the energy consumption using AQNTS-BRF, FDRGS [1] and EERS-CEPO [2]. From the experiment results, the proposed AQNTS-BRF method reduces the energy consumption, thereby improving performance. In this simulation, 10 iterations are performed for measuring the energy consumption during task execution. With minimum numbers of iterations, the AQNTS-BRF method produces better results than other two methods, [1], [2].

When the iterations are increased, the proposed AQNTS-BRF method produces better result than [1] and [2]. The experimental analysis shows that the AQNTS-BRF method consumes minimum energy with multi-objective into considerations task scheduling. Through the above experimental results, it can be observed that AQNTS-BRF can significantly meet the requirements of resource and can save much more time upon comparison to other methods. Moreover, the AQNTS-BRF is found to be enhanced and balanced on exploration and exploitation via Secant Quasi-Newton model and has better scalability. Thus, the AQNTS-BRF method shows its efficiency to improve energy efficiency of data center in green cloud computing by 26% compared to [1] and 42% compared to [2].

4.3 Case scenario 2: Scheduling success

Scheduling success is another critical performance metric to evaluate the efficiency of any method performing on green cloud computing. The scheduling success efficiency is compared and demonstrated in table 3 for two distinct numbers of tasks, i.e., 5000 and 10000 respectively. When the number of tasks increased, the scheduling efficiency started to decrease for all methods. This is owing to the reason that with the excessive amount of data received on green cloud computing results in overloading and therefore leads to performance degradation. The proposed AQNTS-BRF has an efficient mechanism for both task scheduling and resource allocation as the tasks are scheduled according to the cloud user requirements and processed according to multi-objective functionality with global scheduling and optimality principle based on their resource availability, which helps to get better results. From the given results in the table, it can be seen that AQNTS-BRF has more scheduling success than [1] and [2].

Table 3 Tabulation for scheduling success

Generations	**Scheduling success (%) – number of tasks = 5000**			**Scheduling success (%) – number of tasks = 10000**		
	AQNTS-BRF	FDRGS	EERS-CEPO	AQNTS-BRF	FDRGS	EERS-CEPO
20	94.7	93.7	92.1	90.5	89.5	84
40	90.25	86.15	82.35	87.35	83.15	82.45
60	88.45	82.35	87.25	84.15	80.75	79.35
80	85.35	79.45	73.55	82	78.45	77
100	82.05	76.35	71.35	80.35	75.35	74.25

From the above table the overall scheduling success using AQNTS-BRF for 5000 numbers of requests were found to be improved by 6% and 9% compared to [1] and [2]. In a similar manner, the overall scheduling success using AQNTS-BRF for 10000 numbers of requests were found to be improved by 4% and 7% respectively. From this inference it is evident that by applying the Elite-based Secant Quasi-Newton function balance between the exploitation and exploration, therefore improving the overall scheduling success.

4.4 Case scenario 3: Makespan

The performance result of the proposed Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) method is checked for makespan. The proposed AQNTS-BRF method is compared with FDRGS [1] and EERS-CEPO [2] for task scheduling. The number of tasks is varied from 5000 to 50000. In this simulation, the tasks are executed 10 times and then the average has been calculated. The average makespan for the executed tasks using proposed AQNTS-BRF, FDRGS [1] and EERS-CEPO [2] are represented in table 4.

Table 4 Tabulation for makespan

Number of tasks	**Makespan (ms)**		
	AQNTS-BRF	FDRGS	EERS-CEPO
5000	6	8.5	12.5
10000	10.30	14.80	20.10
15000	11.10	15.30	20.80
20000	11.80	16.25	21.30
25000	12.00	17.20	21.80
30000	12.60	19.80	24.30
35000	13.10	21.30	28.20
40000	17.50	22.10	30.10
45000	18.05	25.40	32.05
50000	19.80	27.05	34.80

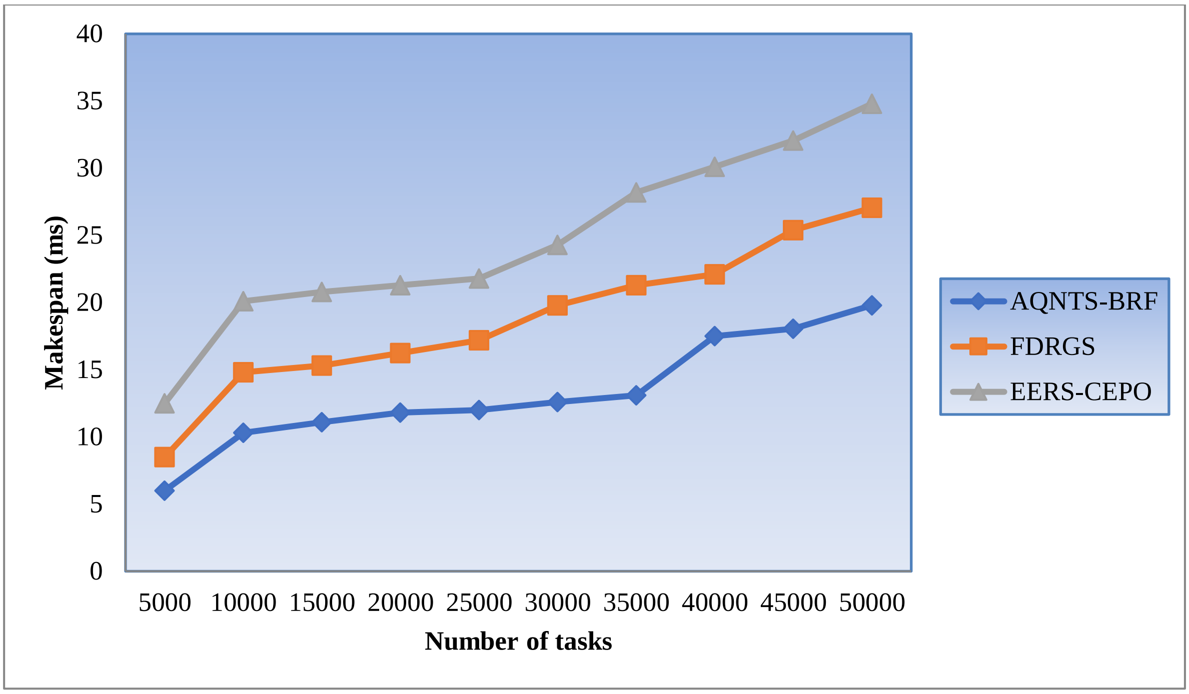


Figure 6 Comparison of the makespan of the four scheduled resource allocation methods

Figure 6 shows the makespan comparison of the three resource allocation algorithms. Here, the y axis represents the makespan and x axis denotes number of tasks. It is clearly evident from the above figure that the proposed HR-SRNN has better performance in terms of makespan upon comparison to other methods [1] and [2]. When the number of tasks is found to be higher in green cloud computing environments, the difference in makespan becomes more apparent. As illustrated in the above figure, when the number of cloud tasks is 5000, the makespan values for AQNTS-BRF, ASA-TL [1] and EERS-CEPO [2] are 6ms, 8.5ms and 12.5ms. Similarly when the number of cloud tasks is 10000, the makespan values for AQNTS-BRF, ASA-TL [1] and EERS-CEPO [2] are 10.30ms, 14.80ms and 20.10ms respectively. Obviously, with the help of the AQNTS-BRF method, task execution time is said to be minimal during the resource allocation process. It can be observed that AQNTS-BRF can obtain optimal solution in an accurate and precise manner. Owing to this also the computation complexity of conventional methods, like [1] and [2] increase the overall performance, therefore compromising the makespan also.

4.5 Case scenario 4: Average response time

To assess the system's efficiency, average response time is measured and their results are presented in table 5. These average response time parameters were compared with AQNTS-BRF, ASA-TL [1] and EERS-CEPO [2]. During the task scheduling and consecutively the resource allocation process in green cloud computing a significant amount of response time is said to be consumed. This is the ratio of total service time to waiting time. When the number of tasks increased (i.e., from 5000 to 10000), the task scheduling efficiency and consecutively the resource allocation started to decrease for all methods. This is owing to the reason that the excessive amount of data received on cloud server results to overload them and leads to performance degradation. However, the proposed AQNTS-BRF has an efficient mechanism for task scheduling as well as for resource allocation employing Bellman Reinforcement Learning–based Fair Resource Allocation as the tasks are scheduled and processed based on their requirements and resource availability that in turn aids in getting better results.

Table 5 Tabulation for average response time

Generations	**Average response time (ms) – number of tasks = 5000**			**Average response time (ms) – number of tasks = 10000**		
	AQNTS-BRF	FDRGS	EERS-CEPO	AQNTS-BRF	FDRGS	EERS-CEPO
20	28.35	35.25	41.35	38.30	45.30	51.40
40	38.15	49.35	58.15	48.10	59.40	68.20
60	45.25	55.15	69.25	55.30	65.20	79.30
80	51.35	68.35	75.35	61.40	78.40	85.40
100	60	71.45	80.25	70.05	81.50	90.30

From the above table the overall average response time using AQNTS-BRF for 5000 numbers of requests were found to be improved by 20% and 32% compared to [1] and [2]. In a similar manner, the average response time using AQNTS-BRF for 10000 numbers of requests were found to be improved by 17% and 27% respectively. From this inference it is evident that by applying the transitional probability matrix and reward functions via Bellman Fair Reward function optimality is arrived, therefore reducing the average response time also.

5. Conclusion

Reduction in energy consumption of Cloud infrastructure is prerequisite to ensure sustainable and green Cloud computing. As several tasks are deployed to the Cloud, maintenance of their QoS requirements (shorter makespan, reduce response time, improving scheduling success) is also necessary. In this paper, we have proposed an energy efficient task scheduling and makespan aware resource allocation called, Antlion Quasi Newton Task Scheduling and Bellman Reinforcement Fair (AQNTS-BRF) in green cloud computing environment. Our first algorithm Antlion and Quasi Newton Multi-objective-based Task Scheduling is based on a multi-objective approach and schedules tasks using Elite-based Secant Quasi-Newton function. Our second algorithm Bellman Reinforcement Learning–based Fair Resource Allocation which uses Bellman's principle of optimality, further improves the solution obtained by combining multi-objective learning mechanism (i.e., round trip delay time, CPU utilization and energy consumption) is integrated into Antlion Optimizer. Extensive simulations validate the efficiency of our proposed method in reducing energy consumption, makespan, average response time and improving scheduling success significantly.

References

1) *Rama Rani, Ritu Garg, "Reliability aware green workflow scheduling using ε-fuzzy dominance in cloud", Complex & Intelligent Systems, Springer, Feb 2022 [ε-fuzzy dominance based reliable green workflow scheduling (FDRGS)]*

2) *Romany F. Mansour, Hesham Alhumyani, Sayed Abdel Khalek, Rashid A. Saeed, Deepak Gupta, "Design of cultural emperor penguin optimizer for energy-efficient resource scheduling in green cloud computing environment", Cluster Computing, Springer, Jan 2022 [Energy Efficient Resource Scheduling using Cultural Emperor Penguin Optimizer (CEPO) algorithm, called EERS-CEPO]*

3) *Mohammed Joda Usman, Abdul Samad Ismail, Hassan Chizari, Gaddafi Abdul-Salaam, Ali Muhammad Usman, Abdulsalam Yau Gital, Omprakash Kaiwartya, Ahmed Aliyu, "Energy-efficient Virtual Machine Allocation Technique Using Flower Pollination Algorithm in Cloud Datacenter: A Panacea to Green Computing", Journal of Bionic Engineering, Springer, Nov 2019*

4) *Anees Ur Rehman, Zulfiqar Ahmad, Ali Imran Jehangiri, Mohammed Alaa Ala Anzy, Mohamed Othman, Arif Iqbal Umar, Jamil Ahmad, "Dynamic Energy Efficient Resource Allocation Strategy for Load Balancing in Fog Environment", IEEE Access, Nov 2020*

5) Juan, Xiaolu Xu, "EERA: An Energy-Efficient Resource Allocation Strategy for Mobile Cloud Workflows", IEEE Access, Nov 2020

6) Harvinder Singh, Anshu Bhasin, Parag Ravikant Kaveri, "QRAS: efficient resource allocation for task scheduling in cloud Computing", Springer, Oct 2021

7) Thi Ha Ly Dinh, Megumi Kaneko, Ellen Hidemi Fukuda, and Lila Boukhatem, "Energy Efficient Resource Allocation Optimization in Fog Radio Access Networks With Outdated Channel Knowledge", IEEE Transactions on Green Communications and Networking, Vol. 5, No. 1, Mar 2021

8) Zhou Zhou, Mohammad Shojafar, Mamoun Alazab, Jemal Abawajy, and Fangmin Li, "AFED-EF: An Energy-Efficient VM Allocation Algorithm for IoT Applications in a Cloud Data Center", IEEE Transactions on Green Communications and Networking, Vol. 5, No. 2, June 2021

9) Yu Zhou, Hui Ge, Bowen Ma, Shuhang Zhang and Jiwei Huang, "Collaborative task offloading and resource allocation with hybrid energy supply for UAV-assisted multi-clouds", Journal of Cloud Computing: Advances, Systems and Applications, Springer, Nov 2022

10) İlksen Çağlar and Deniz Turgay Altılar, "Look-ahead energy efficient VM allocation approach for data centers", Journal of Cloud Computing: Advances, Systems and Applications, Springer, Nov 2022

11) Avita Katal, Susheela Dahiya, Tanupriya Choudhury, "Energy efficiency in cloud computing data centers: a survey on software technologies", Cluster Computing, Springer, Aug 2022

12) Harmeet Kaur, Abhineet Anand, "Review and analysis of secure energy efficient resource optimization approaches for virtual machine migration in cloud computing", Measurement: Sensors, Elsevier, Sep 2022

13) Dhaya R., Ujwal U. J., Tripti Sharma, Mr. Prabhdeep Singh, Kanthavel R., Senthamil Selvan, and Daniel Krah, 'Energy-Efficient Resource Allocation and Migration in Private Cloud Data Centre", Wireless Communications and Mobile Computing, Wiley, Feb 2022

14) Ximu Zhang, Min Jia, Xuemai Gu, Qing Guo,, "An Energy Efficient Resource Allocation Scheme Based on Cloud-Computing in H-CRAN", IEEE Internet of Things Journal, Oct 2019

15) R K Jena, "Multi objective Task Scheduling in Cloud Environment Using Nested PSO Framework", Procedia Computer Science, Elsevier, Nov 2015

16) Abdulaziz Alarifi, Kalka Dubey, Mohammed Amoon, Torki Altameem, Fathi E. Abd El-Samie, Ayman Altameem, S. C. Sharma, Aida A. Nasr, "Energy-Efficient Hybrid Framework for Green Cloud Computing", IEEE Access, Jul 2020

17) Mohammad Nour Hindia, Ahmed Wasif Reza, Kamarul Ariffin Noordin, Muhammad Hasibur Rashid Chayon, "A Novel LTE Scheduling Algorithm for Green Technology in Smart Grid", PLOS ONE | DOI:10.1371/journal.pone.0121901 April 1, 2015

18) Syed Hamid Hussain Madni, Muhammad Shafie Abd Latiff, Mohammed Abdullahi, Shafi'i Muhammad Abdulhamid, Mohammed Joda Usman, "Performance comparison of heuristic algorithms for task scheduling in IaaS cloud computing environment", PLOS ONE | https://doi.org/10.1371/journal.pone.0176321 May 3, 2017

19) Zahra Movahedi Nia, Mohammad Reza Khayyambashi, Ali Miri, "A static VM placement and hybrid job scheduling model for green data centers", PLOS ONE | https://doi.org/10.1371/journal.pone.0237238 August 13, 2020

20) Dexian Yang, Jiong Yu, Xusheng Du, Zhenzhen He, Ping Li, "Energy saving strategy of cloud data computing based on convolutional neural network and policy gradient algorithm", PLOS ONE | https://doi.org/10.1371/journal.pone.0279649 December 30, 2022