Energy Cost Savings on Cloud Data Centres using Heuristic Algorithm

Mervin Abey Abraham
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
mervinabey@gmail.com

Dr. Vergin Raja Sarobin M
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
verginraja.m@vit.ac.in

***Abstract*— Cloud computing has become the front face of various development of cloud data centres around the world as it is all in a shared space called the Internet. Cost savings for consumers and broader extensibility of access are some of the provisions offered. Due to the consumption of energy consumed by the cloud data centres, there has been an insignificant resource allocation and migration of virtual machines. This research paper compares various energy aware algorithms and proposes that an optimized or improved Particle Swarm Optimization algorithm would be the best fit resulting in more than 23% savings. The algorithm shifts around VMs based on the workload and places them in the decreasing order of priority. Literature review shows that the PSO, HSA or combined use of these algorithms results slightly better in higher efficiency and low power consumption.**

Keywords— Harmony Search Algorithm,Particle Search Algorithm, cloud computing, low power consumption, task scheduling, Harmony memory.

I. Introduction

Cloud computing can be best explained by the need of storing user's data at a location as the user does not require to have a whole physical storage unit but rather save a big amount of data online as cloud computing involves a transfer, storage and processing of information on the user's side.

The common issues being faced by cloud computing are memory-aware resource management, energy-aware autonomic resource allocation, SLA violation, effective task scheduling and data management. The objective of a cloud computing service is to decrease their cost and provide more effective useful services to their users as they will be getting more popularity among the computing paradigm as to achieve maximum profitability.

The IT industry has been transformed and revolves in cloud computing as its security and physical storage does not have to be maintained by a company as it's all uploaded to data centres. In data centre cloud computing, we can generally classify the infrastructure as a service (IaaS), software as service (SaaS) and services as a platform as a service (PaaS). These services enable great improvement in a business without increasing the corresponding resource requirements [1].

The limitations faced by cloud computing that are to be focused in huge data centres are that of low memory awareness and huge power usage as the carbon emission from them is affecting the global environment and huge cost charges. People want to perform heavy tasks on a computer without having extensive hardware or software memory allocations. Cloud computing takes care of these needs by providing an adaptive architecture for a variety of computers and data repository services which requires virtual machines (VM). As a cloud service provider, it is necessary that they provide maximum profitability by minimizing operating costs and ensuring their migration based on best cluster head selection.

We have 2 major challenges as the workload of the multiple physical machines is within that in machines and if the machines are shut down or put to sleep with a low workload are to consolidate of, Firstly, choosing the best physical machines big enough to have an ICC cricket match in and to have as many physical machines as possible. Secondly, if the number of VMs in-migration will continue to increase as the migration process uses CPU and bandwidth to transfer data from physical machines to virtual machines to its physical machines. Harmony Search Algorithm or Particle Swarm Optimization (PSO) should be able to select the right machine and redistribute and at the same time do number of migrations form and to the physical machine. [2]

II. Literature Survey

Cloud Computing has been growing exponentially in the last decade so have its applications. Many studies continue to explore reducing power consumption by cloud data centres as well as on efficient resource management, task scheduling, etc.

Usage of power-saving approach in virtual systems was first explored during a study conducted on a massive scale resource utilization [3]. The creators applied a software source scalability approach on a combination of hardware scalability and VM consolidation with help of energy management policies at both levels locally and globally. The global policy frees the host with the least workload thereby strengthening coordination of virtual machines which in turn saves energy. The study results depict that their energy-saving technique saves energy utilization by 34%.

On analyzing studies mentioned in [4,5,6,7] as per optimal energy matching algorithm (PABFD) virtual machines are sorted according to the efficiency of their processors in descending order. The machine that performs the sort is then assigned to the node that has the smallest increase in energy.

[8] shows the authors have explored and come up with a flexible scheduling algorithm that can save energy. The processing speed of each node is adjusted based on the task processing speed so that the tasks are completed within the user's time expectations with reduced energy.

When a large number of consolidated systems are in operation, virtualization is the principal solution to having the required separation layer. Virtual machine runtime migration will occur at a great pace between physical hosts in the cluster and would enable VM containers movement depending on the workload. This motion would be complex and dependent on the distribution of physical computing applications. This calls for more analysis on middleware's live device allocation[8].

Scholarly work of a Cloud computing researcher recommended various new approaches to decrease energy costs[9]. They devised a technique or a strategy using the minimum correlation coefficient (MCC) method that will enhance the VM's selected based on CPU usage. The operation of the Virtual Machine is halted when a virtual migration machine is identified thus affecting the host effectiveness.

[10] recommends that the tasks of the system need to be predictable for which a cubic exponential smoothing algorithm (CESCC) can be used to calculate based on the current state and resource distribution. This improves the real-time response of the system along with system stability.

With rising concern on the CO2 footprint and energy costs caused by data centres, runtime placement of applications has become critical. A typical figure to quote is almost 100 Watts per sq. ft which is continuing to increase dramatically as well the huge energy costs, expensive cooling systems and hardware and floor space affect the environment at drastic levels. The idea or the initiative behind the IT sector is to create a "green" or eco-friendly data centre which will optimize the consumption and take account of the hardware resources as well. It is found from this research article that A runtime controller availability is a critical need as the expense incurred is high[10].

A PSO algorithm describes a bird's-like social behaviour. As it says, it has a number of factors and sets that conform to parameters, particulate interaction, momentum velocity, particle velocity. It consists of a number of parameters, iterative parameter adaptation, topography for interacting with particles, accelerating convergence, adapting to additive, binary or integer domains and hybridising with the other algorithms. [10] has detailed out state of the art in PSO that performs cloud computing scheduling activity with low energy consumption.

An innovative research on energy aware strategies for the optimization of virtual machines highlights that most studies sadly regard the physical servers of a virtualized data centre as being homogeneous in nature, although this is an unfounded presumption.[15] The reason is that additional servers are usually connected to a virtualized datacenter to run new facilities, software or new and growing needs, leading to a heterogeneous virtualized data centre ecosystem in data centre operations. Computer settings in a heterogeneous virtualized data centre are always different, and a wide variety of servers resulting in for Eg: in terms of hardware they can be different from single processing to multi-processing units, to memory, and to fixed disk drives, and other components resulting in a possible broad spectrum of server energy consumption characteristics.[14]

The proposed solution was an optimized PSO that could significantly reduce energy consumption by 13-23% [12]. They achieved this by redefining PSO parameters using an inventive encoding scheme in a 2-dimensional grid and an energy aware strategy for a locally optimal result that potentially enhances data analysis consistency. Fully known by a confidence level of 99%, conventional PSO superiority over genetic algorithms was achieved[13]. PSO offers benefits such as faster performance and increased conflict resolution.

[10] has conducted a study on 5 different scheduling algorithms (Genetic Algorithm, FCFS, Max-Min, Round-Robin, and Priority-based job scheduling algorithm) so as to increase resource availability and decrease energy consumption. The study concluded that individually each of these algorithms affects either load balancing, response time, memory storage units etc and recommend deriving a consolidated.

III. Proposed Low Power Consumption Algorithms

While Cloud Computing is becoming the powerhouse of the explosively growing IT sector of the world, due to it's rising carbon footprint and demand for electricity, many-core studies have been done around these concerns. From Fig.1 we can see that previous studies in 2012 had indicated that the power consumption of data centres were about 2% of the global power consumption which is about 300TWh/year and it was estimated that by 2030, we might be looking at a 10 fold increase which could be around 13% of the worldwide electricity consumption [8].

In this section, we shall first introduce the HSA algorithm which focuses on reducing power consumption using smart positioning of VM migration.

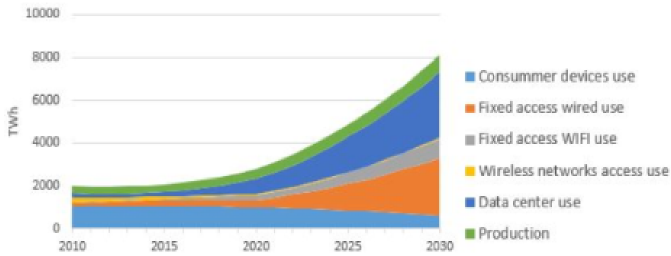


Fig. 1 Power consumption trend of data centres

A. Harmony Search Algorithm (HSA):-

HSA is a meta-heuristic optimization algorithm, inspired by musical improvisation to find the aesthetic form of harmony. To find the optimality in an optimization process like the musician plays a note at a time and continues to perform to form a harmony as harmonic music is analogous in nature. In the algorithm, each note is a variable and during the music, each variable is assigned a value a time and those values all together form harmony and are evaluated by the objective function. In order to understand more about Harmony Search Algorithm (HSA), one must understand that converting the beauty and harmony of music into an optimization process by searching for the perfect harmony is coordinated for idealization into certain quantitative laws in the qualitative improvisation process.

For harmony, there are essentially 2 parameters:- memory and the rate and pitch change rate considered. The right set of cluster heads that minimize carbon emissions needs to be selected. Upon analyzing all of the existing vectors, the HS algorithm introduces a random variable array. It is able to provide a The HS algorithm is also able to help towards them because of the above expertise and functionality. With the approximate residual energy, the sets of cluster heads presented in harmony memory are increased. The high residual energy node is looked for within its clusters in each set of cluster heads. They are expanded by the residual energy calculation and presented in an improved harmony memory [5].

Harmony Memory is a matrix where the number of columns represents dimensions of the result and the rows represent the parameters applicable for each result. This helps in identifying the appropriate variations of the solution that result in efficiency and high performance. A view of the Harmonization Memory Matrix (HM) is shown in the below Fig.2:

$$HM = \left[\begin{array}{cccc|c} x_1^1 & x_2^1 & \cdots & x_N^1 & f(x^1) \\ x_1^2 & x_2^2 & & x_N^2 & f(x^2) \\ \vdots & \vdots & \vdots & \vdots & \\ x_1^{HMS} & x_2^{HMS} & & x_N^{HMS} & f(x^{HMS}) \end{array}\right]$$

Fig 2 HMM

[9] suggests four main stages for VM Migration efficiency:

Stage 1: Sort the physical machines or hosts in descending order on the basis of their task load.

Stage 2: The low-loaded VMs will be selected for migration but depends on their ranking in the migration list.

Stage 3: Add to the target host a VM from the migration list if the host rank is less than the 70% threshold.

Stage 4: Put off the low load hosts. Sample code that increases VM allocation to hosts is shown below:

```
Migration Map optimiseVm Allocation ( Vm List , Host List )
{
   Omit all vms in vm list that in migration or recently assign
    Find all under load hosts ( the model introduced in[34])
    Select vm in under load host to be migrated(the model introduced in[34])
    Placement this vm (Call Hs for this vm)
    For (all under load host) until possible (capacity not full,Ts : 70%)
    Migrate all vms from first host of queue
    Omit host from first of queue power it off
}
```

Fig.3 Pseudo-code

B. Particle Swarm Optimization (PSO)

PSO is a socio-demographic based optimization algorithm created as a reflection of the intellect and strength in animals that always move as a group similar to the pool of fishes.Each fish from the pool will be like a particle of the result with different parameters in consideration such as the velocity a particle maintains and the location of the particle. Locations of the particles are compared against the best fitness value to analyze the effectiveness of the result. PSO begins by initializing particles randomly and then iterates until an optimal result is found. It portrays a similar nature to that of a flock of birds as well i,e. each and every particle moves at a specific speed through the multi-dimensional search space where Xgbest and Xlbesti indicate best position at global and local levels respectively[17].

$$V_i^{t+1} = \omega V_i^t + c_1 r_1 (X_{lbesti}(t) - X_i^t) + c_2 r_2 (X_{gbest}(t) - X_i^t) \quad ,(7)$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} , \quad (8)$$

where V_i^t and V_i^{t+1} are the velocity before the update and the updated velocity, respectively; and X_i^t and X_i^{t+1} are the position before the update and the updated position, respectively. Here, ω is called the inertia weight coefficient, which represents the inheritance of the current velocity of the particle and can balance the local and global search capability of the particles; c_1 and c_2 are called learning factors, which enable the individual to have the ability to learn; and r_1 and r_2 are random numbers that are between 0 and 1.

Now, there are a couple of reasons why this traditional PSO must be optimized. Firstly, a conventional PSO should only be applied for the resolution of the optimization problems that are continuous in nature and not for discontinuous issues to address nonlinear target problems that require a redesign of PSO parameters and operators. Secondly, updating of positioning strategy and coding scheme needs to be updated.

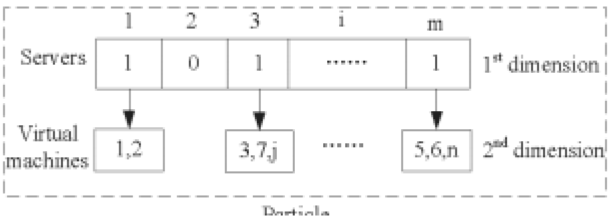
The enhancements as part of PSO Optimization includes 1) Redefining variables and functions that will help solve the positioning aspect 2) adjusting the particle position by using an energy saving strategy that targets fitness primarily.[13]

The redefined parameters and operators are Particle Position, Particle Velocity, Subtraction operator, Addition operator and Multiplication operator. The particle positions are usually adjusted using randomization methods but this may minimize the effectiveness of the model. Using local fitness strategy that will allow the fitness to be in its highest effectiveness as follows:

$$f_{local-i} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \left(\sum_{j=1}^{k} u_{ij}(t)\right) \cdot dt,$$

where uij(t) defines the fitness strategy value and indicates the CPU Utilization of the server .

In order to improve the effectiveness of the solution , two dimensional encoding scheme needs to be as follows:



[17] In comparison to traditional PSO, this type of encoding scheme reduces search time and improves the concurrency velocity. The placement of the virtual machine is best derived when the below steps are followed:

Step A: As an input to the solution, constantly compile all demands of the virtual machine.

Step B: Germinate the process.Initial position and velocity will be derived using the first dimension.

Step C: The local and global best position of every particle is calculated once the fitness of all particles in an initial population is calculated.

Step D: Adjust particle speed.

Step E: Adjust particle position. Updation of local positions can skip virtual machines, but they will be filled in the subsequent steps. There can be a chance where the virtual machines are set up on two or more servers, as a result duplicates need to be removed to sustain solution feasibility. To have a feasible solution, virtual machine backfilling needs to be performed where skipped or removed virtual machines need to be inserted again into the servers. When no active servers are found, a new server will be activated.

Step F: Adjust local and global best particle position considering the updated nw population

Step G: If the present iteration number is lesser than the mentioned maximum iteration number then go to Step C else continue to the next step.

Step H: Obtain best global position and its efficiency in fitness perspective then derive the optimal values for the virtual machine placement.

Step I: Once all the demands from the virtual machines are in the present server the iteration ends.

C. *Cloud data centre energy consumption model*

Energy consumption ratio of the cloud data centre is as follows:

$$S^E = \sum_{i=1}^{N_i} (t_i \times p_i) \Big/ \sum_{j=1}^{N_j} (t_j \times p_j)$$

Among them, S^E represents the energy consumption ratio of the cloud data centre of the two scheduling algorithms. The t_i, p_i and N_i represent the number of activated physical machines, the running time of the numbered i physical machine and power of the first algorithm in the cloud data centre. The t_j , p_j and N_j represent the number of activated physical machines, the running time of the numbered j physical machine and power of the second algorithm in the cloud data centre [1].

IV. Comparative Analysis and Results

A. *K-Means, HSA and PSO Clustering Algorithm Comparisons*

On Fig.3 We can see that PSO hybrid has better optimization than all the other three algorithms. HSA usually does a clustering by finding the harmonic memory and do iterations based on it and find its optimal optimization and K-Means usually tends to cover faster grounds but isn't always as accurate as covering clusters of PSO hybrid as it does better than both K-Means and HSA by seeding the initial swarm with the result in terms of HSA and K-Means algorithm.[16]

The hybrid PSO first executes the K-means algorithm once. In this case the K-means clustering is terminated when (I) the maximum number of iterations is exceeded, or when (2) the average change in centroid vectors is less than 0.0001 (a user-specified parameter). The result of the K-means algorithm is then used as one of the particles, while the rest of the swarm is initialized randomly. The gbest PSO algorithm as presented above is then executed.

B. Results

There are many algorithms available and have been researched for this paper but some of them have pitfalls in other algorithms like PSO, DFS, FCFS, Max-Min, Round-Robin, Priority-based job scheduling algorithm, and Genetic Algorithm.

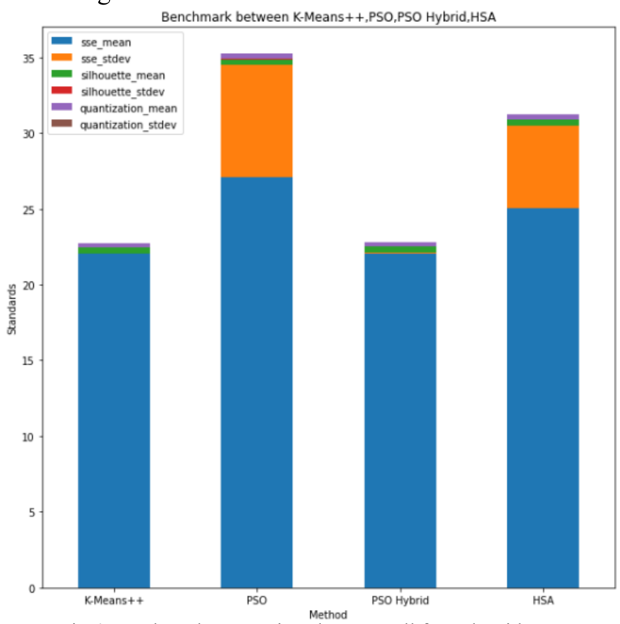


Fig.4 Benchmark comparison between all four algorithms

As PSO faces high dimensional optimization restriction, it is difficult to analyze any possible area of the objective function. In such cases, HSA, which has high computational search capacity. After considering all of the current vectors, it proposes a new way to produce particles that generate a new vector.[13]

In the MCC, to decide if the VM can be transferred and put in other active or reactivated hosts, it is important to locate the load balancing data centre and save it from exhausting any extra power while it is sleeping. The CPU's operating frequency and functioning running voltage to minimize will dynamically change in Dynamic Voltage Scaling (DVS).

In this research I have tried to benchmark HSA, PSO, K-Means++ and Hybrid PSO with benchmarking standards of mean and standard deviation on sum of squared error (SSE), Silhouette and Quantization on all four methods.

On Fig. 4, we are able to note that HSA has its potential with cluster head management during VM but PSO has done better than HSA with a very minute difference which may cost the company a lot for excess power consumption. In this research it was proved that a PSO hybrid can redefine its particle position and velocity of clusters seeds thrown to it in multiple iterations in a matter of seconds and is able to better HSA, PSO and K-Means++.

benchmark

	method	sse_mean	sse_stdev	silhouette_mean	silhouette_stdev	quantization_mean	quantization_stdev
0	K-Means++	22.024363	0.000000	0.422105	0.000000	0.301665	0.000000
1	PSO	27.080251	7.395153	0.395904	0.040273	0.320626	0.011037
2	PSO Hybrid	22.070765	0.016733	0.422127	0.000000	0.301247	0.000147
3	HSA	25.060403	5.423699	0.412855	0.000000	0.321653	0.012078

Fig.4 Benchmark Result

In our approach PSO hybrid was able to solve its optimization problem and was able to adopt the improved / hybrid PSO by reducing the time complexity and clustering its single particle represented by c cluster head centroid vectors.

C. gbest PSO Cluster Algorithm

The PSO cluster of data vectors were done as below:

1. Initialize each particle to contain N_c randomly selected cluster centroids.
2. For $t = 1$ to t_{max} do
   (a) For each particle i do
   (b) For each data vector $\mathbf{z}_p$
      i. calculate the Euclidean distance $d(\mathbf{z}_p, \mathbf{m}_{ij})$ to all cluster centroids C_{ij}
      ii. assign $\mathbf{z}_p$ to cluster C_{ij} such that $d(\mathbf{z}_p, \mathbf{m}_{ij}) = \min_{\forall c=1,\cdots,N_c}\{d(\mathbf{z}_p, \mathbf{m}_{ic})\}$
      iii. calculate the fitness using equation (8)
   (c) Update the global best and local best positions
   (d) Update the cluster centroids using equations (3) and (4).

where t_{max} is the maximum number of iterations.

Where m_{ij} refers to the j-th cluster centroid vector of the i-th particle in cluster C_{ij}. Therefore, a swarm represents a number of candidate clustering for the current data vectors. The fitness of particles is easily measured as the quantization error.where d is defined in equation (I), and C_{ij} is the number of data vectors belonging to cluster C_{ij} i.e. the frequency of that cluster.

Here tmax was its number of samples overall. In contrast to the K-means or HSA algorithm, population-based search of the PSO algorithm limits the impact on initial circumstances. The hunt begins in tandem with many locations. This section introduces a typical gbest PSO to cluster data into a given group and then shows how to merge K-means, PSO algorithms to further boost the PSO algorithm efficiency.

V. CONCLUSION AND FUTURE WORK

Algorithms were selected to target efficient energy consumption with minimal SLA deviation. Low load virtual machines were made inactive and reassigned to physical machines. Before reassignment, factors such as CPU Utilization, memory required for performing cloud tasks are considered for maximum power savings.

In this research we have found out that PSO hybrid does a better job in optimizing clusters and probably would do better virtual machine migrations compared to HSA, KMeans and PS. It is necessary to utilize memory required only for cloud tasks to maximise power savings to avoid SLA violations as much as possible as well.

This research paper shows that applying the PSO hybrid algorithm in cloud data centres could help to save energy cost and bring out-migration efficiency as well compared to other algorithms in this research. PSO - hybrid approach has shown it has better convergence to smaller intra-cluster distance, lower quantization errors and even in larger distances.

One of the directions for future work is to analyze the possibility of other algorithms such as HSA-PSO hybrid Algorithm that can be used combined to achieve higher power energy savings and even resolve other issues like faster response time and To limit energy consumption, HSA can be paired with the PSO algorithm. Use the parallel variant of HSA to minimize cloud reaction times or even to optimize the inter- and intra- cluster distances which will be dynamically extended to proactively calculate the ideal number of clusters.

ACKNOWLEDGEMENT

I would like to thank my professor, Dr. Vergin Raja Sarobin M (Guide) SCOPE, VIT Chennai for guiding me throughout the digital assignment and helping me in various ways as suggesting the changes wherever necessary. I would also like to appreciate Research papers, Quora, Google and all the other authors for providing algorithms, images and documents which were necessarily needed on time.

REFERENCES

[1] Bin Liang, Xiaoshe Dong, Yufei Wang, Xingjun Zhang / Memory-aware resource management algorithm for low-energy cloud data centres (2020)

[2] Anders S. G. Andrae * and Tomas Edler / On Global Electricity Usage of Communication Technology: Trends to 2030 (2015)

[3] R.Nathuji, K. Schwan, "Virtual Power: Coordinated power management in virtualized enterprise systems", ACM SIGOPS Operating Systems Review, vol. 41,p. 265–278, (2007).

[4] A.Beloglazov, J. Abawajy, R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centres for Cloud computing" vol. 28,p.755–768, (2012).

[5] A.Beloglazov, R. Buyya, "Adaptive Threshold Based Approach for Energy Efficient Consolidation of Virtual Machines in Cloud Data Centres",p.1-6, (2010).

[6] A.Beloglazov, R. Buyya, "Energy-Efficient Resource Management in Virtualized Cloud Data Centres",p.826-831, (2010).

[7] A.Beloglazov, R. Buyya,"Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Center", vol. 24,p.1397–1420, (2012).

[8] K. Chen, J. Powers, S. Guo, F. Tian, CRESP: Towards optimal resource provisioning for MapReduce computing in public clouds, IEEE Trans. Parallel Distrib. Syst. 25 (2014) 1403–1412.

[9] Mohammad Alaul Haque Monil and Rashedur M. Rahman / VM consolidation approach based on heuristics, fuzzy logic, and migration control (2016)

[10] Xiaolong Xu, Lingling Cao, and Xinheng Wang / Resource pre-allocation algorithms for low-energy task scheduling of cloud computing (2016)

[11] Esmaile Khorram, Majid Jaberipour / Harmony search algorithm for solving combined heat and power economic dispatch problems (2011)

[12] Shangguang Wang1, Zhipiao Liu1, Zibin Zheng2, Qibo Sun1, Fangchun Yang1 / Particle Swarm Optimization for Energy-Aware Virtual Machine Placement Optimization in Virtualized Data Centers (2013)

[13] E. C. Laskari, K. E. Parsopoulos, and M. N. Vrahatis, "Particle Swarm Optimization for Integer Programming," (2002)

[14] Foudil Abdessamia ,Yu Tai , WeiZhe Zhang , Muhammad Shafiq," An Improved Particle Swarm Optimization For Energy-Efficiency Virtual Machine Placement" (2016)

[15] Shangguang Wang, Ao Zhou, Ching-Hsien Hsu, Xuanyu Xiao, Fangchun Yang, "Provision of Data-intensive Services through Energy- and QoS-aware Virtual Machine Placement in National Cloud Data Centers" (2014)

[16] Shangguang Wang , Zhipiao Liu , Zibin Zheng , Qibo Sun , Fangchun Yang, "Particle Swarm Optimization for Energy-Aware Virtual Machine Placement Optimization in Virtualized Data Centers " (2013)

[17] Vanita Chaudhrani, Pranjalee Acharya and Vipul Chudasama," Energy Aware Computing Resource Allocation Using PSO in Cloud " (2018)

[18] Sadoon Azizi1 • Mazhar Zandsalimi1 • Dawei Li, "An energy-efficient algorithm for virtual machine placement optimization in cloud data center" (2020)