BINARY PROGRAMMING MODELS FOR ENERGY-EFFICIENT VIRTUAL MACHINES PLACEMENT IN DATA CENTERS

Delia Mihaela RADULESCU (NEACSU)[1,2], Marius RADULESCU[3], Constanţa Zoie RADULESCU[4], Gheorghe LAZAROIU[5]

The aim of the paper is to present two virtual machines (VMs) placement models that minimize energy consumption in a data center. Both placement models use binary programming. The first model aims to find an allocation of VMs to a set of servers such that the energy consumption is minimized. The second model suppose that an allocation of VMs to a set of servers is given. Its aim is to find another allocation of VMs to servers that minimize the sum of two energies. The first energy is the energy consumption of VMs in the new allocation and second energy is the energy needed for migration between the two allocations.

Keywords: virtual machines, energy consumption minimization, servers, allocation, data center, binary programming model

1. Introduction

The data center industry had witnessed in the recent years a rapid growth. This is caused by the dramatic increase of the demand for digital services and the number of applications that require computer resources. In the recent years the use of artificial intelligence (AI), internet of things (IoT) and other new technologies boosted electricity consumption of data centers. As a result, the need for energy and water uses in data center facilities knew a rapid increase.

The estimated world data center electricity consumption in the year 2022, cf. [1] was between 240-340 TWh.

[1] Eng., National Institute for Research and Development in Informatics, Bucharest, Romania, e-mail: delia.mihaela2010@gmail.com

[2] PhD student, Faculty of Power Engineering, National University of Science and Technology POLITEHNICA Bucharest, Romania, e-mail: delia.mihaela2010@gmail.com

[3] Senior research worker, Dept. Statistical Inference, Statistical Control and Optimization, *Gheorghe Mihoc-Caius Iacob* Institute of Mathematical Statistics and Applied Mathematics, Romanian Academy, Bucharest, Romania, e-mail: mradulescu.csmro@yahoo.com

[4] Senior research worker, National Institute for Research and Development in Informatics, Bucharest, Romania, e-mail: zoie.radulescu@ici.ro

[5] Prof., Dept. of Energy Production and Use, National University of Science and Technology POLITEHNICA Bucharest, Romania, e-mail: glazaroiu@yahoo.com

This figure does not include energy used for cryptocurrency mining, which was estimated to be around 110 TWh. By 2030, as much as 353 TWh of energy could be used by data centers.

Energy is very important in a data center since it is necessary for running servers, for cooling systems, storage systems, networking equipment, backup systems and security systems. Data centers usually use a lot of energy to maintain active their servers. Their power consumption ranges from a few watts to 2 kW per square meter depending on IT workload.

The greatest amount of energy in a data center is used by servers. A typical server might draw from 100 to 600 watts under standard operation. High-end servers, particularly those running intensive high-performance computing workloads, could consume a few kilowatts of power. The power usage of servers can vary significantly depending on the specific workloads they are executing.

The substantial rise in energy costs and organizations' sustainable initiatives to reduce global warming gave an impulse to academic community for involvement in elaboration of strategies for lowering data center power usage.

There are several strategies for reducing energy consumption of data centers servers. They include:

- Consolidation of under-utilized Servers;
- The use of built-in server power management features. Modern servers come with features that can save energy, but they have to be activated;
- Reducing Energy Losses from Power Distribution Units (PDUs): Procurement of more efficient PDUs and "smart" PDUs that monitor power usage;
- Reducing Energy Losses from Uninterruptable Power Supply (UPS) Systems: Procurement of more energy-efficient UPS systems that minimize electrical losses and have an "eco-mode" feature;
- Managing airflow and cooling efficiency. Using inexpensive grommets, diffusers and blanking panels that can keep cold air from mixing with hot exhaust air. Using Containment/Enclosures: curtains or Plexiglas panels that can keep cold air from mixing with hot air exhausted from the backs of servers. These will mitigate the overall cooling costs.

The aim of the present paper is to present two binary programming models for virtual machines placement on a set of servers from a data center. The VMs placement is made such that the energy consumption be minimized. The structure of the paper is the following. In section 2 a literature review on optimization models in data centers is presented. The third section is dedicated to the server consolidation. In the fourth section a binary programming model for optimal allocation of VMs to servers is formulated. In section 5 an energy minimization model for VMs migration is presented. The paper ends with the conclusion section.

2. Literature review

Resource management in a data center is a difficult decision problem. This fact is due mainly to the heterogeneity of resource types and their interdependencies. The variability and unpredictability of the workload is also a source of difficult problems. In order to solve these problems both academia and industry began to make significant research efforts in this area. In a data center improper utilization of computer resources should be avoided. Reduction of the servers' energy consumption is a main objective of data centers managers. It is motivated by the environmental and economic constraints. The main challenge for data centers managers is to provide services with a low energy consumption without compromising the quality of service (QoS). In the following is presented a literature review on the optimization research papers in the domain of servers' consolidation.

A reference paper in the domain of the static servers' consolidation belongs to Speitkamp and Bichler [2]. The authors had studied the consolidation problem with a binary programming approach. They introduced a cost associated to each server which can be thought as a cost of energy consumption. The objective function in [2] is the total cost (cost of energy consumption).

In [3] a two-level control system is proposed to manage the allocations of workloads to VMs and of VMs to servers. The emphasis is on the VMs placement problem which is formulated as a multi-objective optimization problem. One of the objectives is the minimization of the total resource wastage, energy consumption and thermal dissipation costs. An improved genetic algorithm with fuzzy multi-objective evaluation was proposed for solving the multi-objective optimization problem. Some algorithms for solving the binary programming problems for VMs placement were implemented in [4]. Simulations for checking performances of algorithms were performed with a cloud computing simulation toolkit known as CloudSim using Planet Lab workload data.

In [5] a survey of virtual machine placement techniques in a cloud data center was presented. In [6] it is shown how the energy consumption of the servers during migration can be estimated. The study is based on the resource utilization parameters and use linear regression techniques. The most significant parameters that influence the energy consumption of the servers during migration were defined. These are: CPU instructions retired, last level cache line misses, and "dirty" pages observed in the source server during migration.

An enhanced levy-based particle swarm optimization algorithm was proposed in [7] for solving the VMs placement problem.

In [8] several binary programming models for the static server consolidation problem were formulated. The models determine the best placement of virtual machines on servers in a data center. One model minimizes the energy consumption of the data center.

In [9] an optimization model that tries to find the best allocation of virtual machines to servers such that the energy consumption is minimized. The model has two objective functions. One objective function is the IT power consumption of the servers. The other objective function measures the energy consumption of the cooling system. The aim of the model is the minimization of the two objective functions.

A rigorous survey and comparisons of the binary programming methods for VMs placement in data centers was presented in [10]. Various methods were discussed and the VMs placement factors in each method were analyzed to understand the advantages and drawbacks of each method. However, the server overload/underload detection algorithms were not considered.

In [11] the optimal solution for the problem of VMs placement is computed by using a hybrid optimization method with fitness parameters.

A comprehensive and systematic survey of existing energy-efficient techniques in data center server consolidation was presented in [12]. The survey includes limitations of techniques and the challenges associated in their implementation.

3. Server consolidation

Data centers consume a large amount of electricity, which may generate carbon emissions and environmental pollution. That is why saving the power consumption in a data center is a primary importance task for data centers managers. Reducing energy consumption of servers is very important for saving energy of a data center. It can be done by using a technique called server consolidation. An application of this technique has as a result the reduction of the number of active servers in the data center. Also, this technique allows the reduction of the number of underutilized servers that may draw more energy that is necessary for processing their workload.

Server consolidation is based on a technology called virtualization that improves utilization efficiency of servers and reduces energy consumption. As a result of virtualization several virtual servers (called virtual machines) are built that can run their own operating systems independently. Virtualization is made with the help of software applications called hypervisors. Organizations use server virtualization software to allocate server resources among virtual machines in order to optimize workload management. The virtualization software also increases physical server utilization, reduces costs, reduces the time for deployment, makes easier the system management, improves security and disaster recovery and enhances flexibility.

Virtualization is a growing trend in the last years. Today it can now be considered an indispensable tool that is used regularly by organizations in large-scale server consolidation projects with IT service providers.

Some examples of software products available on the market that enable server virtualization are:

- VMware vSphere which is best for large enterprises and cloud providers.
- Microsoft Azure Virtual Machines: Best for Microsoft integration.
- Citrix Hypervisor: Best for virtual desktop and application deployment.
- Red Hat Virtualization: Best Linux-based virtualization solution.
- Oracle VM VirtualBox: Best high-performance solution.
- Parallels: It offers a range of solutions including Parallels Desktop for Mac, which allows users to run Windows on any Mac—Intel or Apple silicon.
- Xen: It is an open-source type-1 or bare-metal hypervisor, which makes it possible to run many instances of an operating system or different operating systems in parallel on a single machine.

In order to see how servers' energy consumption could be saved we shall present the notion of idle server. Such a server is an underutilized server that is it does not deliver information or computing server when it is on. One can say that it is in a state of inactivity or dormancy. The existence of idle servers is a source of energy waste because they are not efficient.

For example, the power of a non ENERGY STAR® five-year-old server could be around 175 watts when it is idle._A study on the server utilization cf. [13] found that around of 30% of data center servers are in the state of 'comatose'. This means that they haven't produced any useful work within the last six months. There are cases when an idle server can draw about 66% of the maximum power drain. As a conclusion the management of idle servers is essential for energy efficiency in data centers.

In order to give a range for the idle server energy consumption the firm Vertiv had proposed a standard called "the 10 Minus server idle energy standard". This standard had established a benchmark equal to 10 percent of full rated power when the server is idle. Also, it had introduced several ratings for the power consumption servers in the idle state. The Vertiv's ratings are:

- Silver: For those servers whose power consumption in the idle state is below 8.5% of maximum power drain.
- Gold: For those below 7.5%.

- Platinum: For those below 5%.

This standard helps data center managers to make procurement decisions that reduce the energy consumption of idle servers. As a result, the energy efficiency of data centers is increased.

During periods when the number of workloads is small one can reduce the number of active servers (servers in the state on) by migrating VMs to a smaller number of servers and turning off idle servers. This procedure is called consolidation. It helps to the reduction of data center energy consumption as well as of the running costs. The consolidation has its drawbacks. One consequence is the degradation of services' performance during VMs migration. Another consequence is that the migration process requires additional resources such as CPU cycles and network bandwidth. Note that energy consumption increases as a result of the increase of CPU activity. As a conclusion VMs migration consumes energy.

4. Formulation of a binary programming model for optimal allocation of VMs to servers

Consider a data center with *m* servers (physical machines). Suppose that on each server are available *p* resources. CPU, memory, bandwidth etc. are examples of server resources. Denote by $S = \{S_1, S_2, ..., S_m\}$ the set of servers and by $R = \{R_1, R_2, ..., R_p\}$ the set of resources. By $VM = \{VM_1, VM_2, ..., VM_n\}$ we denote the set of virtual machines in the data center.

The parameters of the models are presented in the following.

c_j = the power needed for the server S_j when it is in the idle state, that is when it is on and has no workload.

a_{jk} = the power needed for the use of one unit of resource R_k from server S_j

d_{ik} = the amount (number of units) of resource R_k needed by the virtual machine VM_i

b_{jk} = the amount of resource R_k available on server S_j

The decision variables are:

$$x_{ij} = \begin{cases} 1 & \text{if } VM_i \text{ is active on server } S_j \\ 0 & \text{if } VM_i \text{ is not located on server } S_j \end{cases}$$

$$y_j = \begin{cases} 1 & \text{if server } S_j \text{ is on} \\ 0 & \text{if server } S_j \text{ is off} \end{cases}$$

Let $\mathbf{x} = (x_{ij})$ and $\mathbf{y} = (y_1, y_2, ..., y_m)$.

Note that:

- $C_1(\mathbf{x}) = \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{jk} d_{ik} x_{ij}$ is the power needed for processing the IT workload,

- $C_2(\mathbf{y}) = \sum_{j=1}^{m} c_j y_j$ is the power needed for the idle servers, that is for servers that are on and are without workload,
- $C(\mathbf{x}, \mathbf{y}) = C_1(\mathbf{x}) + C_2(\mathbf{y})$ is the overall power needed for processing the IT workload of the data center for an allocation (**x, y**) of virtual machines to servers.

In the following we shall formulate two models for optimal placement of virtual machines in a data center. The first model searches for a VMs placement in a data center such that the power consumption is minimized. The second model is an energy minimization model for VMs migration. It looks for an allocation that minimize the sum of two energies $\mathbf{E}_1$ and $\mathbf{E}_2$. $\mathbf{E}_1$ is the energy needed for VMs migration from an initial allocation to an optimal allocation and $\mathbf{E}_2$ is the energy needed for VMs placement in the optimal allocation.

The power consumption minimization model

The aim of this model is to find an optimal allocation of virtual machines to servers such that the overall power consumption of the data center be minimized.

$$\begin{cases} \min[C(\mathbf{x}, \mathbf{y})] \\ \sum_{j=1}^{m} x_{ij} = 1, \quad i \in \{1,2,\dots,n\} \\ \sum_{i=1}^{n} d_{ik} x_{ij} \le b_{jk}, \quad j \in \{1,2,\dots,m\},\ k \in \{1,2,\dots,p\} \\ x_{ij} \le y_j,\ i \in \{1,2,\dots,n\},\ j \in \{1,2,\dots,m\} \\ x_{ij}, y_j \in \{0,1\},\ i \in \{1,2,\dots,n\},\ j \in \{1,2,\dots,m\} \end{cases} \tag{1}$$

The constraint $\sum_{j=1}^{m} x_{ij} = 1$ shows that the virtual machine VM_i is placed on a server and this server is unique. The constraints $\sum_{i=1}^{n} d_{ik} x_{ij} \le b_{jk}$ show that the maximum available resources on each server are limited. The constraints $x_{ij} \le y_j$ show that if the server S_j is off that is if $y_j = 0$ then no virtual machine may be located on S_j.

In the case we sum the constraint $\sum_{i=1}^{n} d_{ik} x_{ij} \le b_{jk}$ over j we obtain:

$$\sum_{i=1}^{n} d_{ik} \left(\sum_{j=1}^{m} x_{ij} \right) \le \sum_{j=1}^{m} b_{jk} \tag{2}$$

hence:

$$\sum_{i=1}^{n} d_{ik} \le \sum_{j=1}^{m} b_{jk} \tag{3}$$

The above condition is a necessary condition for the existence of a solution of the power consumption minimization model.

There are situations when it is necessary that a set of VMs be placed on the same server. Such situations occur when VMs from the given set frequently communicate between them. This can help to reduce network latency and improve the performance of the applications running on these virtual machines. For example, if you have a web server and a database server that are often communicating, it might be beneficial to collocate them on the same server. This is one of the strategies used in managing resources in a virtualized environment.

Suppose that K is a subset of $\{1, 2, \ldots, m\}$ with at least two elements, that is $|K| \geq 2$. Let $(VM_i)_{i\in K}$ be the set of VMs that should be placed on the same server. In order to introduce the above constraint in our model we shall introduce a set $(z_j)_{1\leq j\leq m}$ of binary variables. We shall add to the constraints of our model the following set of constraints:

$$\begin{cases} \sum_{j=1}^{m} z_j = 1, \quad i \in \{1,2,\ldots,n\} \\ \frac{1}{|K|}\sum_{i\in K}^{n} x_{ij} \leq z_j, \quad j \in \{1,2,\ldots,m\} \\ \\ z_j \in \{0,1\},\ j \in \{1,2,\ldots,m\} \end{cases} \tag{4}$$

The case when all the servers are identical worth to be investigated.

If all the servers are identical then:

$a_{1k} = a_{2k} = \cdots = a_{mk} = a_k$, $c_1 = c_2 = \cdots = c_m = c$.

Note that in this case $C_1(\mathbf{x})$ is constant:

$$C_1(\mathbf{x}) = \sum_{k=1}^{p} a_k\left[\sum_{i=1}^{n} d_{ik}\left(\sum_{j=1}^{m} x_{ij}\right)\right] = \sum_{k=1}^{p}\sum_{i=1}^{n} a_k d_{ik} \tag{5}$$

Since:

$$C(\mathbf{x},\mathbf{y}) = \sum_{k=1}^{p}\sum_{i=1}^{n} a_k d_{ik} + c\left(\sum_{j=1}^{m} y_j\right) \tag{6}$$

it follows that the minimization of $C(\mathbf{x},\mathbf{y})$ is equivalent with the minimization of the function $N(\mathbf{y}) = \sum_{j=1}^{m} y_j$.

Note that the minimization of $N(\mathbf{y})$ is equivalent with the minimization of the number of active servers.

The above optimization model is a much more complete version of the model presented in [8]. The necessary condition for the existence of a solution of the minimization model and the VMs collocation on the server constraint on the same server are new contributions.

5. Energy minimization model for VMs migration

This model is an energy minimization model for VMs migration. It looks for an allocation that minimize the sum of two energies $\mathbf{E}_1$ and $\mathbf{E}_2$. $\mathbf{E}_1$ is the energy needed for VMs migration from an initial allocation to an optimal allocation and $\mathbf{E}_2$ is the energy needed for VMs placement in the optimal allocation.

Let T be the length of a time interval. T could be an estimation for the length of the time interval between two VMs migrations.

Then the VMs energy consumption in the allocation ($\mathbf{x}$, $\mathbf{y}$) during the time interval is:

$$\mathbf{E}_2(\mathbf{x},\mathbf{y}) = T \cdot C(\mathbf{x},\mathbf{y}). \tag{7}$$

Consider that it is given an allocation $\mathbf{x}^{(0)} = \left(x_{ij}^{(0)}\right)$ of VMs to servers. We want to find an allocation $\mathbf{x} = \left(x_{ij}\right)$ of VMs to servers such that the energy consumption of VMs that is $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ is minimized.

For every $i \in \{1,2,\dots,n\}$ denote:

$$g_i(\mathbf{x}) = \sum_{j=1}^{m} x_{ij}\, x_{ij}^{(0)} \tag{8}$$

Note that $g_i(\mathbf{x})$= 1 if VM_i does not move in the migration process and $g_i(\mathbf{x})$=0 if VM_i moves on another server in the migration process. One can easily see that:

$$g(\mathbf{x}) = \sum_{i=1}^{n} g_i(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}\, x_{ij}^{(0)} \tag{9}$$

is the number of VMs that do not move on another server in the migration process.

Note that $g_i(\mathbf{x})$= 1 if VM_i does not move in the migration process and $g_i(\mathbf{x})$=0 if VM_i moves on another server in the migration process.

Note that:

$$1-g_i(\mathbf{x}) = \begin{cases} 0 & \text{if } VM_i \text{ does not move in the migration process} \\ 1 & \text{if } VM_i \text{ moves on another server in the migration process} \end{cases}$$

Denote by E_i the energy consumption of VM_i migration from one server to another server. E_i depends on the amount of resources needed for VM_i

Let $\mathbf{E}_1(\mathbf{x})$ be the energy needed for VMs migration from allocation $\mathbf{x}^{(0)} = \left(x_{ij}^{(0)}\right)$ to allocation $\mathbf{x} = \left(x_{ij}\right)$. Then $\mathbf{E}_1(\mathbf{x})=\sum_{i=1}^{n}\left(1 - g_i(\mathbf{x})\right)E_i$ is the energy consumption of all VMs migration from allocation $\mathbf{x}^{(0)} = \left(x_{ij}^{(0)}\right)$ to allocation $\mathbf{x} = \left(x_{ij}\right)$.

The minimization of energy consumption model for VMs migration is the following:

$$\begin{cases} \min\left[T \cdot C(\mathbf{x},\mathbf{y}) + \sum_{i=1}^{n}\left(1 - g_i(\mathbf{x})\right)E_i\right] \\ \sum_{j=1}^{m} x_{ij} = 1, \quad i \in \{1,2,\dots,n\} \\ \sum_{i=1}^{n} d_{ik}x_{ij} \le b_{jk}, \quad j \in \{1,2,\dots,m\},\ k \in \{1,2,\dots,p\} \\ \quad x_{ij} \le y_j,\ i \in \{1,2,\dots,n\},\ j \in \{1,2,\dots,m\} \\ \\ x_{ij}, y_j \in \{0,1\},\ i \in \{1,2,\dots,n\},\ j \in \{1,2,\dots,m\} \end{cases} \tag{10}$$

Parameter T could be used by the data center manager for managing the trade-off between the energy consumption of the virtual machines in the new allocation and the energy needed for migration between the two allocations.

The above optimization model is original.

Binary programming models can be solved with various methods proposed by many researchers. It is known that the binary programming problems are NP-Hard. There are several commercial software libraries that contain solvers for large scale binary programming problems: GAMS, Gurobi, HIGS, SCIP etc. They can handle thousands of decision variables and constraints. However, for solving large scale binary programming problems several heuristic procedures were developed.

For solving virtual machines placement and migration problems in the literature were proposed heuristic algorithms based on several simple methods such as: next fit (NFD), first-fit decreasing method (FFD) and best-fit decreasing method (BFD). For an approach based on hybrid heuristic algorithms see [14]-[20].

6. Conclusions

Energy consumption of data centers depends on the proper resource utilization. If the resources management is not proper (over-utilization or under-utilization) then the energy efficiency consumption in the data center will be low. As a result, the profit of the data centers managers will decrease drastically. As a conclusion, energy-aware server consolidation becomes a well-known issue in the field of data center management. Server consolidation using virtualization technology has become an important method for improving data center energy efficiency. This technique enables physical servers to host multiple independent virtual machines, and the workloads migration from one server to another.

The aim of the paper is to present two virtual machines placement models on servers that minimize energy consumption in a data center. Both placement models are linear binary programming models. The objective of the first model is to find an allocation of virtual machines to a set of servers such that the energy consumption is minimized. The special case when the servers are identical is studied. Also, the introduction of supplementary constraints for sets of virtual machines that need to communicate between them is made. Necessary conditions for existence of solutions of the binary model are formulated.

The second model suppose that an allocation of virtual machines to a set of servers is given. Its aim is to find another allocation of virtual machines to servers that minimize the sum of two energies. The first energy is the energy consumption of virtual machines in the new allocation and second energy is the energy needed for migration between the two allocations.

In the second model a parameter T occurs. It is connected with the time the initial placement of virtual machines to servers is used. However, parameter T could be used by the data center manager for managing the trade-off between the energy consumption of the virtual machines in the new allocation and the energy needed for migration between the two allocations. Future work will consider heuristic algorithms for solving binary programming models proposed in this paper.

REFERENCES

[1]. *Statista*, https://www.statista.com/statistics/186992/global-derived-electricity-consumption-in-data-centers-and-telecoms/

[2]. *B. Speitkamp and M. Bichler*, "A mathematical programming approach for server consolidation problems in virtualized data centers", IEEE Trans. Serv. Comput., **vol. 3**, no. 4, 2010, pp. 266–278.

[3]. *J. Xu and J. A. Fortes*, "Multi-objective virtual machine placement in virtualized data center environments." In 2010 IEEE/ACM international conference on green computing and communications and international conference on cyber, physical and social computing, December 2010, pp. 179-188. IEEE.

[4]. *M. R. Chowdhury, M. R. Mahmud and R. M. Rahman*, "Study and performance analysis of various VM placement strategies." In 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), June 2015, pp. 1-6. IEEE.

[5]. *Z. Usmani and S. Singh*, "A survey of virtual machine placement techniques in a cloud data center." Procedia Computer Science, *78*, 2016, pp 491-498.

[6]. *K. Rybina and A. Schill*, "Estimating energy consumption during live migration of virtual machines." In 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), June 2016, pp. 1-5. IEEE.

[7]. *A. Fatima, N. Javaid, T. Sultana, W. Hussain, M. Bilal, S. Shabbir and M. Ilahi, "*Virtual machine placement via bin packing in cloud data centers." *Electronics*, 2018, *7*(12), 389.

[8]. *D. M. Rădulescu, C. Z. Rădulescu, G. Lazaroiu and R. Boncea*, "Binary programming models for the server consolidation problem in data centers." International Multidisciplinary Scientific GeoConference: SGEM, 2018, *18*(2.1), pp 201-208.

[9]. *D. M. Rădulescu, G. Lazaroiu and C. Z. Rădulescu, "*Power consumption minimization models in data centers." International Multidisciplinary Scientific GeoConference: SGEM, 19(2.1), 2019, pp 517-524.

[10]. *S. Kumaraswamy and M. K. Nair, "*Bin packing algorithms for virtual machine placement in cloud computing: a review." International Journal of Electrical and Computer Engineering, 9(1), 2019, 512.

[11]. *S. Supreeth, S. Bhargavi, R. Margam, H. Annaiah and R. Nandalike, "*Virtual Machine Placement Using Adam White Shark Optimization Algorithm in Cloud Computing." SN Computer Science, *5*(1), 2023, 21.

[12]. *N. Chaurasia, M. Kumar, R. Chaudhry and O. P. Verma, "*Comprehensive survey on energy-aware server consolidation techniques in cloud computing." The Journal of Supercomputing, 77, 2021, 11682-11737.

[13]. Time for a Server Idle Performance Standard – Introducing 10 Minus Vertiv, https://www.vertiv.com/en-asia/about/news-and-insights/articles/blog-posts/time-for-a-server-idle-performance-standard/

[14]. *M. A. Kaaouache and S. Bouamama*, Solving bin packing problem with a hybrid genetic algorithm for VM placement in cloud. Procedia Computer Science, *60*, 2015, pp. 1061-1069.

[15]. *H. Belbachir, S. Bouroubi and A. Khelladi*, "Connection between ordinary multinomials, Fibonacci numbers, Bell polynomials and discrete uniform distribution" Annales Mathematicae et Informaticae Vol. 35, 2008, pp. 21-30.

[16]. *M.A. Păun, H.G. Coandă, E. Mincă, O. G. Duca, I. Făgărășan, F. Dragomir and S.S. Iliescu, "*Advanced multi-objective algorithm used to optimize consumption of an integrated system for flexible manufacturing" U.P.B. Sci. Bull., Series C, Vol. 86, Iss. 2, 2024, pp. 113-126.

[17]. *F. Guerriero and F. P. Saccomanno*, "A hierarchical hyper-heuristic for the bin packing problem" Soft Computing, *27*(18), 2023, 12997-13010.
[18]. *V. Romero, S. Osaba, E. Villar-Rodriguez, E. Oregi, and Y. Ban,* "Hybrid approach for solving real-world bin packing problem instances using quantum annealers" Scientific Reports, *13*(1), 2023, 11777.
[19]. *Z. Mahmoodabadi and M. Nouri-Baygi*, "An approximation algorithm for virtual machine placement in cloud data centers" The Journal of Supercomputing, 80(1), 2024, 915-941.
[20]. *M. C. Çavdar, I. Korpeoglu and Ö. Ulusoy,* "A Utilization Based Genetic Algorithm for virtual machine placement in cloud systems" Computer Communications, *214*, 2024, 136-148.