A Review on Energy Efficient Approaches for Cloud Computing

Riakshi Routray[1*] Swagatika Behera[2*] Namrata Dash[3*] R P R Terdeage Murmu[4*] Ashis Kumar Mishra[5*]
Computer Science and Engineering
Odisha University of Technology and Research (OUTR) Bhubaneswar, India

Abstract:- Cloud computing is becoming widely adopted, which increases both its user's efficiency and the provider's profitability. Virtualization is assisting the cloud in increasing resource utilization, but it is also increasing operational expenses due to increased energy usage. This paper studies and sheds light on the contemporary modern day research on cloud energy management techniques such as ballooning, virtual machine migration, resource allocation (RA), EEEHVMC (Energy Efficiency Heuristic with Virtual Machine Consolidation), three-way decision (VMM-3WD), and others that can be implemented in fundamental software level, virtualization and various applications and the operating system itself. It studies various measures to reduce the energy usage and clearly contributes to minimizing existing environmental problems and curbing it. This paper clearly explains each approach, including its benefits, drawbacks, and future prospects, taking into account modern day parameters. This article also shed light on the burdensome parameters and criteria's that cloud organizations and cloud data centers must follow.

Keywords:- *Cloud computing, Virtualization, VM migration, SLA violation, Load prediction, Task Scheduling.*

I. INTRODUCTION

Cloud computing is a virtualization of processors, memory, computation, networking and storage technology combined into a platform that enables dynamic scaling and fast application deployment. One of the biggest advantages of cloud computing is the pay-as-you-go policy, that charges users on the basis of the resources they use. In order to provide well adaptable resources, fast easy access and scale economies, cloud computing simply refers to the storehouse of computing services—including storage, software, servers, networking, databases, intelligence and analytics over the cloud. Technologies for automation and virtualization are crucial to cloud computing. As a result of automation and corresponding orchestration features, users can link services, provide resources and run workloads achieving a rewarding level of self-service, without having to involve the cloud provider's IT professionals' support. Cloud computing is gaining popularity among clients owing to the numerous benefits it provides at a reasonable cost. Energy usage has raised the expense of cloud data centers due to a growth in the number of clients, thereby increasing the load on customers' pockets. Data center energy demand will increase to 2967 TWh in 2030 than 200 TWh in 2016. To provide services, data centers expend a lot of energy, which results in increased carbon dioxide emissions. In this survey report we examine software-based solutions, which incorporate power control at the individual programmer level, that might be utilized to create green data centers. The paper investigated container problem-solving energy efficient methodologies for lowering power usage in data centers. Furthermore, the study discusses the environmental effect of data centers, including e-waste, as well as the varied criteria used by different nations to rate data centers. This article shows more than simply new green cloud computing opportunities. Instead, this work concentrates society's and academia's resources on a significant problem: long-term on-going technologies trend. The study discusses emerging methods that may be used at the basic software part, including approaches used at the operating system, virtualization and application levels. It outlines many strategies for cutting energy use at every level, adding on to the present issue of reducing environmental pollution. The paper also discusses some of the variables and case studies that affect the use of green clouds, as well as the challenges, worries, and requirements that cloud service providers and cloud enterprises must understand. The energy utilization of various data centers only, will hike to 2967 TWh in 2030 than 200 TWh in 2016[1]. Instead of the COVID-19 issue, the global market for Internet and Data Centers is anticipated to hike at a CAGR of 13.4% over 2020 and 2027, which was before estimated US$59.3 billion in 2020 to US$143.4 billion by 2027 [2]. By 2020, the Internet Data Centers market in the US is anticipated to be worth US$16 billion. With a 17.5% CAGR from 2020 to 2027, China, the second-largest economy in the world, is anticipated to get a data center market of US$32 billion that year [2]. Virtualization, which offers several virtual instances on restricted physical machines/servers, is the key approach for decreasing energy use. Cloud computing has grown in popularity as a new technology that helps the enterprises to employ computational capabilities such as application, software, hardware and so on to perform computing through the internet. This computing theory enables the use of cloud infrastructure at many levels of thought. Cloud computing has received a lot of light as it provides enhanced accessibility and adaptability at a cheap cost. Cloud computing provides storage and compute capacity for usage by third-party services. Cloud computing is utilized by various applications (apps) [3]. Cloud computational services are scalable reliable and available on a pay-as-you-go basis. Relying entirely on cloud resources has some limitations[3]. This paper examines and sheds light on contemporary cloud energy management techniques such as ballooning, virtual machine migration, resource allocation (RA), EEEHVMC (Energy Efficiency Heuristic with Virtual Machine Consolidation), three-way decision

(VMM-3WD), and others. All of these characteristics include CPU usage, makespan, cost, migrations, and service level agreement breaches, execution. This work briefly describes each approach, including its benefits, drawbacks, and future prospects. The major objectives of this study are to present a systematic examination of cloud infrastructures by developing a detailed research and nomenclature for energy efficiency and to give a summary of the technologies, research, and projects that are now in use.

II. PRELIMINARY STUDIES AND BACKGROUND

The authors of [4] investigated the various migration operations in OpenStack in the presence of severe system limits and networking difficulties. They conducted a migration performance study, which might aid academics in developing better optimizing strategies. The authors of [5] gave a summary of pre-existing live migration strategies, including the hypervisor name, performance metrics, basic approach (post-copy, pre-copy, and combination of both) employed, and any benefits and downsides. The authors of [6] proposed a taxonomy of live virtual machine migration, which gives information of "scheduling and planning algorithms", "cost and performance models", "migration generation in resource management rules", "orchestration and management lifecycle", and "evaluation techniques". The study also discussed and examined the present status of "migration scheduling kinds", "migration network awareness", "scheduling scope", "homogeneous and heterogeneous solutions", and "scheduling objectives**".** Liu et al. [7] gave a method for VM consolidation in cloud computing context. The key idea was to minimize the number of VMMs and hence the amount of energy utilized in an environment like the cloud. The suggested technique achieves the following major goals: lowering the probability of host overload, preventing undesired VMMs, and, as a result, minimizing the overall number of VMMs. The suggested strategy enhances host computer resource usage. It performs admirably under various workload scenarios. As a result, this technique meets the criteria of reducing data center costs for resource suppliers. Prospects for this technology are advantageous for both providers that provide services and end customers. According to Uddin et al. [8], a server consolidation approach can boost the effectiveness of pre-existing machines and their usefulness by converting them to virtualized machines in order to support environmentally as well as economically friendly cloud data centers. Tasks from real server machines are evenly distributed to virtual machines using an unique virtualized work scheduling technique. The outcomes of the experiment showed that 30% might improve the effectiveness of resource use (in the VMs). Furthermore, the study conducted found that using the fewest servers (up to 50%) resulted in considerable energy savings. The EVMC "Energy-Aware VM Consolidation" system [9] employs a resource and parameter-based strategy to manage overutilized hosts in the virtual cloud context. Comparisons of VMs and hosts are used to identify overcrowded hosts, whereas the CATR "Cumulative Available-to-Total Ratio" is used to identify underused hosts, transfer VMs to appropriate hosts; VM placement employs a criteria based on normalized host and virtual machine resource metrics. To determine the performance of VM consolidation, several tests were run on several virtual machines utilizing traces from PlanetLab workloads. The results indicate that the EVMC strategy performs similarly to other well-known strategies in terms of energy savings, SLA breaches, and the frequency of VM migrations. Recent research [10] suggests the PSO "Particle Swarm Optimization" model for cost reduction and scheduling with deadline constraint as a prominent example of the metaheuristic method. Though they highlight the benefits of cloud computing, like the utilization of unlimited distinct resources, versatility in terms of obtaining and, heterogeneity, releasing resources and the pay per use model, major challenges such as energy efficiency and CO2 emissions are not addressed. [11] gives a proper solution that combines various methods for resource efficient allocation and proper task scheduling. To enhance response time, a mix of BATS that is bandwidth-aware task scheduling and BAR is used in assigning jobs. Since the system only takes into account bandwidth availability when preemption, allocating resources in this stage, and divide and conquer procedures are used to execute tasks if the virtual machine is overburdened with the job that was allocated in the BATS+BAR step. The absence of using metaheuristic strategies to address energy inefficiency is highlighted as an unresolved problem. In [12], which compares metaheuristic scheduling algorithms and analytic findings. However, more steps are required to execute workloads since processes to check resource availability are ignored, which results in energy overhead. Only a few studies in another large review on task scheduling systems [13] examine energy usage while offering task scheduling using metaheuristic approaches. Consolidation is a common strategy used in allocation of resources and migration to conserve energy by using fewer resources and increasing resource utilization rate. Consolidation can be accomplished using heuristic procedures both during allocation and migration. Using heuristic methods that are rapid enough for large-scale systems but don't guarantee optimal results, the migration allocation and placement operations are treated as a Bin Packing optimization [14]. In the primitive ABC-artificial-bee colony algorithm, frustration from the aimlessly chosen bee in a random evaluation is used to produce the food supply. In the ABC method, failure machines are picked at random, which does not assist achieve optimal energy utilization. The author created a VM migration strategy based on the ABC-Bat algorithm [15] to increase energy efficiency. The suggested method is assessed using a variety of criteria, including energy use [15], success rate accuracy as well. Furthermore, it has performed well in terms of 1-1.2kWh minimum energy usage [15], 0.2 with maximum completion rate [15], and accuracy of 1 and 97.77%. The humpback whales' poaching behaviour inspired the whale optimization algorithm (WOA) [16]. WOA has three mechanisms that mimic the search for prey, including the bubble net invasion process and the prey. Whale optimization algorithm was ferocious and superior than

ISSN No:-2456-2165

previous approaches; multi-objective WOA is now being developed. ChicWhale optimization method [18] is used to choose the appropriate VM by combining the chicken swarm optimization algorithm with WOA. In this technique, a provincial migration operator is used to administer memory and resource utilization on a regular basis, and migration is executed by the service provider depending on the VMs' requirement for resources to complete an assigned job. The ChicWhale Algorithm supports the load significance [18] of each VM housed in the cloud and conducts a check by applying a certain threshold; if the load exceeds this threshold, the work is reassigned to another VM. Following that, the suggested approach demonstrates the new fitness function in terms of several parameters and reassigns the jobs/tasks to appropriate VM in order to complete the fitness computation. Qu et al. [19] represented the DMRO, that uses reinforcement machine learning for offload workloads from edge-cloud servers and IoT devices etc. Karthiban and Raj [20] employed same reinforcement learning to perform balancing of load in CDC, despite the complicated network topology between users and resources. Sun et al. [21] presented a reinforcement learning strategy for offloading a computational operation in vehicle cloud and edge cloud. Ma and Ding [22] presented a reinforcement learning-based methodology for determining which system variables are associated with energy usage in a CDC. To minimize SLAV and energy usage, Joseph et al. [23] suggested using RL approaches to handle the challenge of properly mapping a provided set of microservice containers to the resources set. Bitsakos et al. [24] presented DERP (a reinforcement learning-based technique) that by default adjusts diverse computer resources based on user demands, providing cloud elasticity. Cheng et al. [25] introduced DRL-cloud, that employs the RL approach to allocate a huge number of jobs to multiple servers, lowering energy usage while assuring that works are finished before the deadline. Electricity price changes are also established as limitations in DRL-cloud to ensure that energy expenses are fully taken into consideration. Tuli et al. [26] suggested a scheduling technique based on reinforcement machine learning and Recurrent neural networks for scheduling of application workloads between cloud and edge servers in order to improve SLAV and energy utilization.

III. MODERN METHODOLOGY

In [27], it employs the ballooning approach, which deletes unnecessary data prior to transferring the VM. The approach conserves time as compared to other strategies since it reduces the amount of information that has to be sent. The task is accomplished by counting the data accesses such as memory disc block or memory page, performed while the virtual machine executes, and any low valued data are targeted for deletion. The current ballooning method is unable to distinguish between a page that was just generated and has no accesses and a page that was created many years ago and has the least number of accesses. As a result, in this technique, time taken for generating a single piece of data is also considered, such that newly produced data is of substantial relevance and not deemed as unnecessary. A framework that makes advantage of virtual memory is suggested in [28]. When the host runs out of memory, ballooning is utilized to collect unused memory from virtual machines. A balloon driver running on a guest VM retrieves the memory. Ballooning of memory is manually activated such that the specified virtual machine's unused, free, and inconsequential pages are recovered, reducing the expense associated with moving those pages from source to destination. Other solutions include pre-copying the VM's memory after ballooning at the time the execution remains with the source host. The precopy is terminated at a moment known as the Time Switch Decision, which is the point at which the precopy is terminated and the postcopy is initiated. The execution state and the remainder of pages in memory are transmitted in the Post-Copy technique. After migrating the memory and execution states, the Virtual Machine will be entirely moved to the target host. The process of migration is complete, and the VM begins operating on the target Host. The initial step in [29] is to identify hosts depending on their load condition as underloaded hosts, overloaded hosts or regular load hosts. Then, several migration procedures are devised with these three categories of cloud hosts in mind. The method, in particular, migrating VMs from under-utilized servers to typical load hosts. The technique then produces two criteria to categorize overloaded hosts as massively, moderately or lightly overloaded. The decision of VM migration at all stages with the intent of minimizing network energy usage throughout the migration procedure. The function of resource scheduler is to offer alternatives for resource allocation on the basis of characteristics of resources like demand statistics, resource monitoring and use and the public cloud in [30]. When a request arises, schedulers provide either stable and preliminary allocation of resources or both stationary and dynamic resource allocation. To decrease overall energy consumption in cloud data centers, the migration approach is paired with an algorithm that results in accurate allocation. The precise allocation technique, when combined with migrations, significantly reduces the number of extra cloud host servers required to manage a known stack, minimizing power consumption in cloud services. The main components of Energy Efficient Scheduler are migration and allocation. The objective of these elements is to finish the preliminary virtual machine deployment using a precise virtual machine allocation approach. Because of our accurate virtual machine migration algorithm, the migration component controls the active consolidating in VMs, lowering the number of utilized or host servers that are active. Total power consumption is calculated by adding idle power consumption and CPU use, then multiplying peak power by idle energy consumption. Using dynamic VM consolidation to minimize power usage and SLA breaches in cloud data centers [31], Based on two thresholds, hosts are categorized as: HML “Host Medium-Loaded”, HOL “Host Over-Loaded” and HUL “Host Under-Loaded”. To reduce energy usage, the proposed approach reallocates virtual computers from one PM to another. The EEHVMC technique minimizes power consumption along with SLAV by using the memory and CPU. To reduce power usage in cloud data centers, we relocate VMs to the HML for HOL.

All VMs in the HML are left alone. According to the suggested method, hosts that are inactive are switched into power-saver mode and the VMs are transferred from the HUL hosts to the HML. The total power consumed by a physical machine is the sum of the server's constant power consumption irrespective of whether VMs are operational or not while the dynamic power consumed by the VMs operating on the server. For migration, the VM that incurs the highest CPU utilization is chosen. The VM having the highest memory consumption is chosen first from among those that have been chosen. Transferring a big quantity of memory may require more resources, but because the memory source of the host is limited, the slow rate of migration will almost certainly result in additional delays. When CPU-intensive operations overburden a host, it proposes a novel VM selection mechanism called Maximum ratio of CPU usage to memory utilization (MRCU) to pick migrating VMs. The total complexity is O (number of host computers x number of virtual machines).In [32],proposed algorithm optimizes job distribution to VM, which reduces the frequency of virtual machine migrations, cutting cost of migration, and determining the appropriate virtual machine for migration, which aids in handling tasks efficiently, controlling the load among physical hosts, and eventually reducing energy usage. Originally, the cloud framework was established, and tasks were optimized with the aid of a load balancer in the cloud data centers and given to VMs, reducing the number of virtual machine migrations and, as a result, the migration cost. The proposed technique in [33] makes use of predictive methodology. Because of its adaptability to time series, Holt Winters is favored as a forecasting tool. It is presented as a resource management method for virtualized data centers that works on optimizing the number of servers to satisfy the demands of changing workloads without VM migration. LAA is compared against CloudSim's best virtual machine migration technique, dubbed LR-MMT" Local Regression-Minimum Migration Time ". LAA allows for the adaptive assignment of incoming user queries to computer resources. A specific threshold is used to identify overutilization of CPU, but it isn't the only criteria considered when deciding whether to start a new server that can host the newly incoming task. During the allocation choice process, the trend of the system is equally significant as the threshold. If the number of servers that are currently active fulfills the anticipated future demand along with the present overutilization scenario happens briefly, then the proposed method allocates the new workload by taking into account the time remaining for workloads that are already operating on active servers. The suggested approach is made up of the following functional modules: monitoring, workload allocation and forecasting. The function of the monitoring module is to keep track of the execution time (ET) and CPU requirement (CR) of each inbound task for the previous 5 minutes. The monitoring module sends the information it has gathered to the forecasting and task placement modules. The forecasting module calculates the necessary number of processing units (Nr) based on demand by the user. The workload placement module is in charge of assigning workloads to appropriate VMs and assigning these VMs to appropriate servers. Algorithm 1: works on the system's diminishing or growing trend. Enlarging is the default tendency in which the system will grow to fulfill new workloads by adding new servers. Because of the declining trend, the system seems to have more operational servers than is necessary for anticipated demand. Algorithm 2: The server which has the most time for the remainder of the workloads is chosen. Algorithm 3: Run the workload which has the least execution time after assignment, based on the ideal utilization rate, to shorten the time the chosen server is operating when it is overloaded. The authors in [4] suggest that using MiGrror instead of pre-copy can result in decreased migration time, downtime as well as delay. MiGrror is a novel migration technique and procedure that is based on VM/container mirroring. When switching between source and destination virtual machines or containers, mirroring is utilized to synchronize them more rapidly. At the conclusion of a round that represents a certain period of time, pre-copy communicates dirty memory. MiGrror focuses on minimizing the time needed to transmit the last of dirty memory ($t_{transfer}$) of the virtual machine/container since the objective is to limit the quantity of data that has to be transmitted at downtime to gain improved performance. The difference between the two techniques is that MiGrror synchronizes (syncs) the source and the destination while the events happen, as opposed to pre-copy, which waits for a round to conclude. Every memory update at the source generates an event, signaling that both the source and destination has to be synchronized. MiGrror does not have to wait for anything to happen. MiGrror, on the other hand, allows for numerous correspondence of the destination and source throughout the pre-round copies in order to replicate the current virtual machine/container accessible to the destination. MiGrror synchronized at the very instant a memory change takes place and transmits differences in memory when they are accessible. When the hand-off is initiated, a little quantity of dirty memory is still present. This is the data in the final memory variation provided by the source to the destination to synchronize after the hand-off trigger. Because other memory variations have already been transferred, this round of pre-copy has less memory than the previous round. As a consequence, $t_{transfer}$ in MiGrror is lowered when compared to pre-copy. As a result, downtime is decreased when compared to pre-copy. Algorithm 1 describes the synchronized pseudo-code for migration of the source edge node, from whence migration begins. Algorithm 2 shows the destination edge node's synchronized migration pseudo-code, which is where migration ends. In [34], the author offers a VM selection technique that is learning-based that chooses suitable virtual machines for migration without involving direct overloading detection of the host, minimizing the creation of SLAV, assuring performance, and lowering cloud data center energy consumption. Overloading is not always associated with SLA violation, but it should be present when that occurs. The goal of screening out overloaded hosts is to relocate certain VMs ahead of time in order to prevent formation of SLAV or to get virtual machines out of SLAV status. As a result, rather than overwhelming the hosts, the purpose of virtual machine migration may be configured to prevent the occurrence of SLA violation. Our suggested technique detects the hosts that may become overloaded in the near future, as well as which VMs should be transferred, in order to optimize energy usage and reduce SLAV. Reinforcement learning is used to remove the host overload detection function. The time of migration is computed by distributing the VM's memory use by the available bandwidth. The MMT technique chooses the VM that requires the least amount of time to migrate and adds it to the list of VMs to migrate. The technique first detects if the host is overloaded by counting the historical CPU resource utilization by the given host using the local regression method. A random host is chosen from the list of hosts to be the destination to migrate the current virtual machine. FF chooses the first accessible host from the host list to be the destination for the migration of the current VM. In [35], proposed an energy-efficient algorithm for VM allocation that optimizes execution of tasks within a secure virtualized environment by extending the Intelligent Water Drop (IWD) algorithm to the WDVMA (Water Drop Virtual Machine Allocation). It identifies over- and under-loaded hosts and migrates VMs accordingly to enhance server usage. In addition, offer a secure cloud architecture for data privacy and protection. For encryption, the Advanced Encryption Standard (AES) algorithm is employed. The IWD method is algorithm 1, and the proposed WDVMA algorithm is algorithm 2. The first stage in determining the best approach is to convert the VM Allocation issue into the IWD method and generate a fully connected graph G (V, S) with n number of nodes connected by edges. Each node in the network represents a virtual computer, and there are exactly as many nodes as there are virtual machines. Initialize each edge of the graph G in a two-dimensional array by giving it a constant value. After the initial iteration, each of the nodes is traversed depending on that edge value, and an array list is generated to keep the record of visiting nodes. We kept the array list for visited nodes and the edge updated. The node is chosen based on that edge value; the edge with the lowest value is chosen to be the next step for the path. If the values of two edges are the same, choose the edge that has not been visited and consumes the fewest resources. After each iteration, a new step is noted, and this step provides a result for that iteration. The current result is compared to the prior result, which if proved to be better, the current solution is kept and the next iteration is used to identify the ideal solution.

A. Publication Statics

The statistics of publications in different methodologies of the past 10 years (from the year 2010 to 2020) are reported graphically in figure 1.

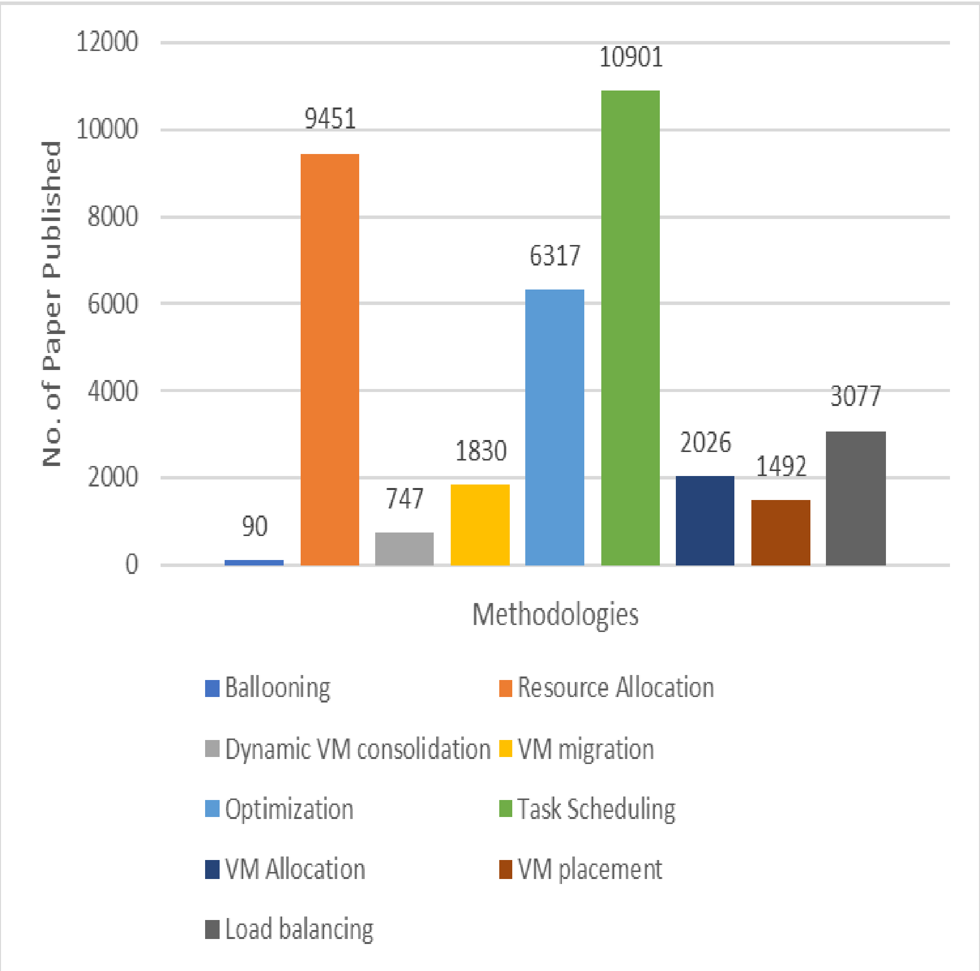

Fig. 1: Published articles trend in different modern methodologies

The diagrammatic representation of the number of published papers by multiple sources is shown in figure 2.

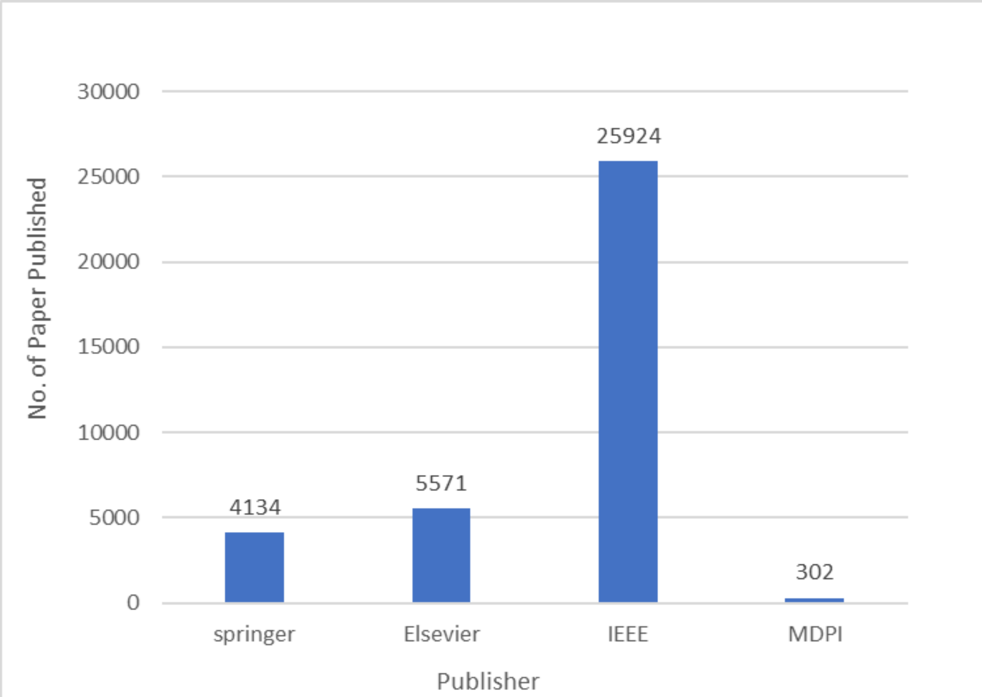

Fig. 2: Classification of published articles over the last 10 years

B. Taxonomy On Energy Parameters

Energy consumption- The entire energy consumption in cloud data centers might boost client costs. The major steps in this respect are virtualization and live VM migration. If conducted often, live VM migration might have a negative impact on the system.

The total energy consumption (operational server) = fixed energy consumption + dynamic energy consumption

- Migration time- It specifies the time interval that elapses between the stage when a VM starts migrating and its resumption on the destination server.
- Makespan- It is defined as the measurement of the total time taken for processing a set of tasks to its completion.
- SLAV (Service level agreement violations)- It is a metric to determine what percentage of legally agreed upon regulations of using cloud services has been obliged upon.
- Throughput- It is the measure of the number of units of information a system can process within a specified interval of time.
- Resource utilization- It is the metric that measures how efficiently a particular resource is utilized in the data center.
- CPU utilization- It describes the percentage of CPU instances distributed on the basis of tasks backed by priority, user-initiated tasks or system-initiated tasks.
- Memory utilization- It is defined as the percentage of average utilization out of the amount of available memory at any instant.
- Downtime- It is defined as the interval of time when the virtual machine remains non-operational so that its migration to the destination server can be carried out.
- Response time- It is defined as the time taken to initiate the process of migration of a virtual machine after the decision for its migration has been taken.
- Atomicity- It is the property whereby migration of a virtual machine is carried out independent of the execution of other VMs.
- Convergence- It is defined as the situation when the storage state and memory between the source and the destination server are practically equal. Convergence implies that the copying of data has been successful.
- Cost and security- Minimizing the power and energy usage during VM migration and providing security to sensitive data is of utmost importance.
- CO2 emission- The process of VM migration heats up the servers which result in functioning of cooling systems along with the consequential emission of CO2.

IV. COMPARATIVE ANALYSIS OF LATEST TRENDS

Table 1 describes the details about the existent research and work in the field of energy management and optimization in cloud with their pros and cons.

Table 1: Review Of Various Methodologies

Author(s)	**Methodology**	**Advantages**	**Limitations/ Future scope**	**Simulation**
Neha et al. [27] (2022)	Proposed LMEB (Live migration with efficient ballooning), a new live VM migration algorithm. To lower the overall consumption of energy, LMEB focuses on reducing the data size that must be transferred from the source to the destination server.	1. Lowers energy consumption by 39%, downtime by 25% and migration time by 46%.	1. Unable to reduce total time for which servers remain functional (Makespan of servers).	JAVA
Saima et al. [26] (2019)	Proposed a migration framework that makes use of ballooning before the migration process begins, removing the overhead of sending unnecessary pages to the destination host. Following ballooning, the VM is moved using a hybrid technique that begins with precopy and ends with postcopy.	1. Reduces the number of pages stored in memory 2. Cuts down memory and resource utilization along with the page faults 3. Reduction of overall migration time	…	Not specified
Jiang et al. [29] (2021)	Provided a VM migration strategy developed on a decision that goes three ways (VMM-3WD) to reduce the usage of energy in cloud data centers while also considering correlation of networks among virtual machines	1. Minimizes the energy usage and the instances of VM migration 2. Ensures that the SLAV is kept to a minimum 3. Reduce total energy consumption.	1. The three-way decision concept will be used to solve the issues and bottlenecks of VMs that arise with technology involving migration as other such areas are examined at a more granular level.	CloudSim
Dhaya et al. [30] (2022)	Used a quadratic mathematical optimization strategy. This work uses three algorithms ,1st algorithm aims at efficient resource allocation 2nd algorithm detects overloaded hosts using a linear numerical program (LNP) algorithm and 3rd detects under-filled VMs prevalent within a private cloud.	1. Reduce migration time by 28% 2. Reduce 58% VM downtime	1. Overall number of migrations is large 2. Can be reduced using ml	JAVA
Arshad et al. [31] (2022)	It presents a method called EEHVMC "energy efficiency heuristic using virtual machine consolidation". It aims at minimizing the energy consumption in the cloud. This approach is based on reallocating VMs from one physical machine to another to reduce energy usage. The EEHVMC method uses memory and CPU to lower SLAV and power consumption.	1. Reduce energy consumption, VM migration, SLA Violation 2. Least performance degradation 3. Execution time is reduced	1. The heuristic approach does not provide an optimal solution. 2. EEHVMC does not follow Hadoop or Spark models which are used in CC.	CloudSim
Shalu et al. [32] (2022)	Proposed R&D-WOA (a Bio-inspired Re-initialization and Decomposition-Whale Optimization Algorithm) by combining NSGA-II with WOA to formulate Pareto-optimal solutions. Optimizing task distribution to virtual machines which reduces the number for VM migrations, which in turn lowers migration costs and, ultimately, lowers energy usage.	1. Proposed algorithm results in improving migration cost and energy usage by 67%, 7%, and 1% respectively than the existing WOA.	1. Do not improve resource Abundance	CloudSim

Deniz et al. [33] (2022)	Proposed an approach, called LAA “Look-ahead Energy Efficient VM Allocation,” which follows a prediction module called Holt Winters. It uses forecasting techniques. Help to reduce energy consumption and improve throughput and system performance.	1. Reduce energy consumption 2. Improve throughput and system performance 3. Avoid unnecessary turning on or off of servers	1.Focuses on CPU intensive workloads	CloudSim
Arshin et al. [4] (2022)	Propose a new migration method called MiGrror for fog/edge computing to minimize downtime. Using MiGrror rather than pre-copy could achieve reduced migration time, downtime, and delay, which results in improved performance.	1. Low downtime 2. Less data loss 3. Low average latency 4. Low migration time	1. Use more network bandwidth	MobFogSim
Huixi et al. [34] (2022)	Proposed a VM selection approach that is learning based which decides suitable virtual machines to migrate without involving direct overloading detection of hosts. SLAV generation is decreased, performance is maintained, and the energy usage is decreased.	1.The SLAV generated significantly less 2. Reduce generation of SLAV 3. Reduce energy consumption.	1. The energy usage of VMs will significantly rise if the resource demands for the VMs rise dramatically. Energy consumption rises because additional hosts are needed to load these VMs in order to lower SLAV 2. Need to improve the convergence performance	OpenAI
Kalka et al. [35] (2022)	Proposed an energy efficient virtual machine allocation method which optimizes the execution of tasks in the cloud by extending the IWD(Intelligent Water Drop) algorithm to WDVMA(Water Drop Virtual Machine Allocation) algorithm.	1. Less number of VM transfer 2. Reduce energy consumption and resource utilization	1. Does not consider other QoS parameters 2. Does not consider computation cost and makespan	CloudSim
Moses et al. [36] (2022)	Proposed an optimized technique for resource allocation by load balancing technique implementing a machine learning algorithm. The LSTM algorithm examines heuristic application resource usage to choose the optimal resources that can be supplied with that particular application.	Increases network efficiency by utilizing less memory and CPU time than existing models and having better predictive abilities.	1.The method was unable to determine minimization of power in wired and wireless networks. 2. The approach's conclusions cannot be generalized until it is evaluated with additional network data.	JAVA
Liang et al. [37] (2021)	Proposed PFLBS “a port-based forwarding load balancing scheduling” solution for Fat-tree based DCNs. First a source-routing addressing system based on ports has been devised. Second, an effective routing system for flow scheduling was formulated. Third, an efficient technique for dynamically scheduling big flows based on current link use ratio was tested.	1.Can successfully manage all flows (small and big) while minimizing overhead and lowering switch complexity thanks to the multipath selection mechanism and large flow scheduling approach. 2.Decreases average flow finishing time, results in improved throughput of big flows, and fully utilizes network capacity.	The data-mining effort is more difficult to manage in terms of load balancing.	OMNET++

Malik et al. [38] (2021)	Proposed an approach based on work categorization and criteria for effective scheduling and improved resource use. Preprocessing of workflow tasks are carried out to minimize bottlenecks in the initial phase by registering those tasks with maximum dependencies and longer execution time. Then the tasks are grouped on the basis of intensity of the required resources. The final step involves choosing optimum schedules by implementing Particle Swarm Optimization (PSO).	The suggested approach was validated by experimental findings in terms of energy efficiency, load balancing and makespan. A percentage improvement increases of 13 to 53% was realized.	1. The concerns of virtual machine migration and adaptive thresholds need to be investigated further in order to make the system efficient and obtain better outcomes. 2. Implementing the simulated method in real-world contexts will provide additional obstacles, such as administrative costs, energy spent by causes other than computation, hardware failures, and data backups.	CloudSim
Ramzan et al. [39] (2023)	The suggested COTD approach is divided into three stages. The reaction time is used to calculate the threshold in the first phase. In the second step, the deadline for the task is compared to the minimum reaction time of the accessible data centers. The final step includes comparing the task deadline to the estimated threshold. Finally, based on the deadline, the requests are routed to an appropriate data center.	1.The algorithm's low complexity makes it appropriate for runtime decision making. 2. The suggested method successfully reduces costs by 35% on average whereas keeping response time constant.	To determine the algorithm's efficiency, more complicated scenarios should be explored and simulations performed in real-world contexts.	Cloud Analyst
Yadav et al. [40] (2023)	Proposed an improved and extended ordinal optimization strategy to decrease the enormous search space for optimum scheduling within the shortest time possible in order to achieve the goal of lowest makespan.	The suggested optimization approach along with linear regression provide optimum schedules that aid in achieving the shortest possible makespan. Furthermore, using linear regression, the suggested mathematical equation forecasts any future dynamic workload for a minimal makespan time aim.	1.Other issues such as security, efficiency, job prioritization, and energy usage must also be considered to improve overall performance in the cloud environment. 2.This method is applicable to a limited number of virtual machine configurations, cloudlets, and data centers.	CloudSim
Yu et al. [41] (2022)	Proposed a VM consolidation strategy that is both energy efficient and QoS conscious. To anticipate CPU utilization a hybrid model based on prediction built on the grey model combined with the ARIMA model was developed. A methodology for choosing virtual machines (VMs) was presented, known as AUMT, which chooses the VM with the lowest cost when combining average CPU use and migration time.	When compared to the benchmark technique, the suggested strategy may fulfill the objective of lowering energy usage, the quantity of VM migrations, SLA violations, and ESV	To confirm that the suggested technique is effective in accomplishing the necessary goals, it must be tested on actual cloud platforms.	CloudSim

Kalaivani et al. [42] (2022)	Focuses on maximizing security while decreasing the cost. This paper proposed a Meta heuristic optimization technique for allocating cloud resources. GA-SVM: A novel technique using genetic algorithm (GA) for parameter optimization of SVM (support vector machines) and feature selection.	The applied genetic algorithm produces better processor usage, effective load balancing, make span and cost minimization and optimal throughput	Optimization will be transformed into a more trustworthy and appropriate type of input for the classifier to categorize the many feature categories.	CloudSim and Cloud Analyst
Anand et al. [43] (2021)	Proposed a cost-effective load balancing and resource optimization technique. Implementing various algorithms to properly distribute resources and balance workload in the cloud consumes a lot of energy, and the work is based on the Cuckoo Search Algorithm (CSA), which consumes less energy and takes less time to run than other algorithms.	The proposed study covers a cost-effective method to load balancing and resource planning, achieving significant improvements in resource allocation and reaction time.	The optimization procedure and the addition of hybrid strategies may increase the performance's effectiveness.	Cloud Analyst
Liu et al. [44] (2020)	DCMMT (Dynamic consolidation with migration thrashing) is a proposed approach that puts VM with larger capacity while drastically reducing the number of migrations as well as migration thrashing to achieve SLA.	Improves the SLAV metric by 19%, number of migrations by about 21%, and migration thrashing measure by about 28%.	The research of workload statistical features, such as CDF (cumulative distribution function) and COV(coefficient of variation) is critical for determining the efficacy of this technique.	CloudSim
Taheri et al. [45] (2022)	Proposed a static technique that tries to reduce the execution costs for workflow while still reaching its deadline. The suggested strategy relies on the recurrent usage of the critical route idea to reduce execution costs. The crucial path may be identified in each of the subsequent sub-workflows after breaking up the primary workflow into multiple smaller ones.	EDQWS performs better than its rivals by lowering the cost of execution of out scheduled processes and reaching user-defined deadlines.	Using a non-greedy approach to identify instances of idle time to enable transferring of scheduled activities during the phase of external combination might have yielded better results.	MATLAB

V. CONCLUSION AND FUTURE SCOPE

Cloud computing is performing a crucial role in the computer science and technology industry, because of its enormous quality services provided to the users and end customers. In this article, we systematically analyzed several cloud energy utilization-based methodologies, including problem identification, proposed approaches, advantages, and limits for future work. Due to ecological and economic concerns, there are significant problems in cloud computing for energy optimization. In this regard, we created a comparison chart to help you comprehend the many potential energy consumption strategies. Furthermore, the technique presented by numerous researchers throughout the years has been examined. As future work, we will discover other strategies that give optimal energy usage, as well as the future improvement and practicality of these efforts. Our investigation showed that most typical energy efficient strategies can't perform well for Cloud environment, but must be altered or even built from scratch. This is because the Cloud Processing infrastructure is stratified and consists of devices and parts from several fields of study, including cooling, power supply, automating computing, etc. Separately optimizing these systems improves the overall performance of the system especially in regard to energy. Therefore, if shared energy consumption reduction approaches are used on numerous systems or their parts, the energy efficiency may be greatly increased provided the latest methods are aware of the use case, dependencies and relationships between the systems.

REFERENCES

[1.] Gupta, N., Gupta, K., Qahtani, A. M., Gupta, D., Alharithi, F. S., Singh, A., & Goyal, N. (2022). Energy- Aware Live VM Migration Using Ballooning in Cloud Data Center. *Electronics*, *11*(23), 3932.

[2.] Choudhary, A., Govil, M. C., Singh, G., Awasthi, L. K., Pilli, E. S., & Kapil, D. (2017). A critical survey of live virtual machine migration techniques. *Journal of Cloud Computing*, *6*(1), 1-41.

[3.] Jiang, C., Yang, L., & Shi, R. (2021). An energy-aware virtual machine migration strategy based on three-way decisions. *Energy Reports*, *7*, 8597-8607.

[4.] Sharma, T., Singh, M., Selvan, S., & Krah, D. (2022). Energy-efficient resource allocation and migration in private cloud data center. *Wireless Communications and Mobile Computing*, *2022*.

[5.] Arshad, U., Aleem, M., Srivastava, G., & Lin, J. C. W. (2022). Utilizing power consumption and SLA violations using dynamic VM consolidation in cloud data centers. *Renewable and Sustainable Energy Reviews*, *167*, 112782.

[6.] Singh, S., & Singh, D. (2022). A bio-inspired vm migration using re-initialization and decomposition based-whale optimization. *ICT Express*.

[7.] Çağlar, İ., & Altılar, D. T. (2022). Look-ahead energy efficient VM allocation approach for data centers. *Journal of Cloud Computing*, *11*(1), 1-16.

[8.] Rezazadeh, A., Abednezhad, D., & Lutfiyya, H. (2022). MiGrror: Mitigating Downtime in Mobile Edge Computing, An Extension to Live Migration. *Procedia Computer Science*, *203*, 41-50.

[9.] Li, H., Xiao, Y., & Shen, Y. (2022). Learning-Based Virtual Machine Selection in Cloud Server Consolidation. *Mathematical Problems in Engineering*, *2022*.

[10.] Dubey, K., & Sharma, S. C. (2022). An extended intelligent water drop approach for efficient VM allocation in secure cloud computing framework. *Journal of King Saud University-Computer and Information Sciences*, *34*(7), 3948-3958.

[11.] Ashawa, M., Douglas, O., Osamor, J., & Jackie, R. (2022). Improving cloud efficiency through optimized resource allocation technique for load balancing using LSTM machine learning algorithm. *Journal of Cloud Computing*, *11*(1), 1-17.

[12.] Liu, Z., Zhao, A., & Liang, M. (2021). A port-based forwarding load-balancing scheduling approach for cloud datacenter networks. *Journal of Cloud Computing*, *10*(1), 1-14.

[13.] Malik, N., Sardaraz, M., Tahir, M., Shah, B., Ali, G., & Moreira, F. (2021). Energy-efficient load balancing algorithm for workflow scheduling in cloud data centers using queuing and thresholds. *Applied Sciences*, *11*(13), 5849.

[14.] Ahmad, S. G., Iqbal, T., Munir, E. U., & Ramzan, N. (2023). Cost optimization in cloud environment based on task deadline. *Journal of Cloud Computing*, *12*(1), 1-18.

[15.] Yadav, M., & Mishra, A. (2023). An enhanced ordinal optimization with lower scheduling overhead based novel approach for task scheduling in cloud computing environment. *Journal of Cloud Computing*, *12*(1), 1-14.

[16.] Wang, J., Gu, H., Yu, J., Song, Y., He, X., & Song, Y. (2022). Research on virtual machine consolidation strategy based on combined prediction and energy-aware cloud computing platform. *Journal of Cloud Computing*, *11*(1), 1-18.

[17.] KALAIVANI, R. Energy Efficient and Load Balanced Optimal Resource Allocation framework for Cloud Environment using ML based Meta heuristic techniques

[18.] Anand, B. ENERGY-EFFICIENT SCHEDULING IN CLOUD COMPUTING ENVIRONMENT USING META-HEURISTIC OPTIMISATION

[19.] Liu, X., Wu, J., Sha, G., & Liu, S. (2020). Virtual machine consolidation with minimization of migration thrashing for cloud data centers. *Mathematical Problems in Engineering*, *2020*.

[20.] Khojasteh Toussi, G., Naghibzadeh, M., Abrishami, S., Taheri, H., & Abrishami, H. (2022). EDQWS: an enhanced divide and conquer algorithm for workflow scheduling in cloud. *Journal of Cloud Computing*, *11*(1), 13.

[21.] Gupta, K., & Katiyar, V. (2016). Energy aware virtual machine migration techniques for cloud environment. *International Journal of Computer Applications*, *141*(2), 11-16.

[22.] Chashoo, S. F., & Malhotra, D. (2018, December). VM_Mig_Framework: virtual machine migration with and without ballooning. In *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 368-373). IEEE.

[23.] Gupta, N., Gupta, K., Gupta, D., Juneja, S., Turabieh, H., Dhiman, G., ... & Viriyasitavat, W. (2022). Enhanced virtualization-based dynamic bin-packing optimized energy management solution for heterogeneous clouds. *Mathematical Problems in Engineering*, *2022*.

[24.] Kalra, M., & Singh, S. (2015). A review of metaheuristic scheduling techniques in cloud computing. *Egyptian informatics journal*, *16*(3), 275-295.

[25.] Gupta, K., & Katiyar, V. (2016). Energy aware virtual machine migration techniques for cloud environment. *International Journal of Computer Applications*, *141*(2), 11-16.

[26.] Tuli, K., Kaur, A., & Malhotra, M. (2022, September). Efficient virtual machine migration algorithms for data centers in cloud computing. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 1* (pp. 239-250). Singapore: Springer Nature Singapore.

[27.] Tran, C. H., Bui, T. K., & Pham, T. V. (2022). Virtual machine migration policy for multi-tier application in cloud computing based on Q-learning algorithm. *Computing*, *104*(6), 1285-1306.

[28.] Baker, T., Aldawsari, B., Asim, M., Tawfik, H., Maamar, Z., & Buyya, R. (2018). Cloud-SEnergy: A bin-packing based multi-cloud service broker for energy efficient composition and execution of data-intensive applications. *Sustainable Computing: informatics and systems*, *19*, 242-252.

[29.] Hines, M. R., & Gopalan, K. (2009, March). Post-copy based live virtual machine migration using adaptive pre-paging and dynamic self-ballooning. In *Proceedings of the 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments* (pp. 51-60).

[30.] Wang, X., & Liu, Z. (2012, January). An energy-aware VMs placement algorithm in cloud computing environment. In *2012 Second International Conference on Intelligent System Design and Engineering Application* (pp. 627-630). IEEE.

[31.] Jenitha, V. H. A., & Veeramani, R. (2014). Dynamic memory Allocation using ballooning and virtualization in cloud computing. *IOSR J. Comput. Eng*, *16*, 19-23.

[32.] Sun, G., Liao, D., Anand, V., Zhao, D., & Yu, H. (2016). A new technique for efficient live migration of multiple virtual machines. *Future Generation Computer Systems*, *55*, 74-86.

[33.] Belgacem, A., Beghdad-Bey, K., Nacer, H., & Bouznad, S. (2020). Efficient dynamic resource allocation method for cloud computing environment. *Cluster Computing*, *23*(4), 2871-2889.

[34.] Mazrekaj, A., Nuza, S., Zatriqi, M., & Alimehaj, V. (2019). An overview of virtual machine live migration techniques. *International Journal of Electrical and Computer Engineering*, *9*(5), 4433-4440.

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

[36.] Murtazaev, A., & Oh, S. (2011). Sercon: Server consolidation algorithm using live migration of virtual machines for green computing. *IETE Technical Review*, *28*(3), 212-231.

[37.] Veni, T., Bhanu, S. M. S., & Mary, S. (2013). Dynamic energy management in cloud data centers: a survey.

[38.] International Journal on Cloud Computing: Services and Architecture, 3(4), 13-26.

[39.] Panwar, S. S., Rauthan, M. M. S., & Barthwal, V. (2022). A systematic review on effective energy utilization management strategies in cloud data centers. *Journal of Cloud Computing*, *11*(1), 1-29.

[40.] Deshmukh, P. P., & Amdani, S. Y. (2018, August). Virtual Memory Optimization Techniques in Cloud Computing. In *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)* (pp. 1-4). IEEE.

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

[42.] Kaur, J., & Chana, I. (2018, December). Review of live virtual machine migration techniques in cloud computing. In *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)* (pp. 1-6). IEEE.

[43.] Mohiuddin, I., Almogren, A., Al Qurishi, M., Hassan, M. M., Al Rassan, I., & Fortino, G. (2019). Secure distributed adaptive bin packing algorithm for cloud storage. *Future Generation Computer Systems*, *90*, 307-316.

[44.] He, T., & Buyya, R. (2021). A Taxonomy of Live Migration Management in Cloud Computing. *arXiv preprint arXiv:2112.02593*.

[45.] Sharma, O., & Saini, H. (2021). Performance evaluation of VM placement using classical bin packing and genetic algorithm for cloud environment. In *Research Anthology on Multi-Industry Uses of Genetic Programming and Algorithms* (pp. 1456-1470). IGI Global.

[46.] Aldahwan, N. S., & Ramzan, M. S. (2022). Descriptive literature review and classification of community cloud computing research. *Scientific Programming*, *2022*, 1-12.

[47.] Beloglazov, A., & Buyya, R. (2012). Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurrency and Computation: Practice and Experience, 24(13), 1397-1420.