KeAi
CHINESE ROOTS
GLOBAL IMPACT

International Journal of Innovation Studies

A SURVERY OF OPTIMIZATION TECHNIQUES FOR ENERGY-EFFICIENT CLOUD CONTAINER SCHEDULING

Ms. A. Janaki Devi
Research Scholar
Department of Computer Science
Dr.M.G.R.Educational and Research Institute
Chennai, India
janakidevi08@gmail.com

Dr. V J Chakravarthy
Professor
Faculty of Computer Applications
Dr.M.G.R.Educational and Research Institue Chennai, India.
chakravarthy.mca@drmgrdu.ac.in

Dr Sangeetha Varadhan
Assistant Professor,
Faculty of Computer Applications
Dr.M.G.R.Educational and Researh Institute Chennai, India.
sngth.sangeetha@gmail.com

Dr. N Jayashri
Assistant Professor,Research Scholar
Department of Computer Applications,
Dr.M.G.R.Educational and Research Institute
Chennai, India
jayashrichandrasekar@yahoo.co.in

Dr. V Dheepa
Assistant Professor
Department of Computer Science
Arulmigu Kapaleeswarar Arts and Science College
Chennai, India
dsvdeepasaro@gmail.com

Abstract

Light virtualization techniques known as cloud computing containers have shown to be promising in providing cloud services. This is because it is scalable, portable and simply decorated in a variety of ways, especially for cloud computing MicroD services. Due to the adoptable properties of the cloud content nature of cloud containers, some of the schedulers are extremely important for improving performance and reducing costs. This study classifies planning strategies for four groups based on optimization methods that use schedules, mathematical linear approaches, heuristic approaches, metaheuristic approaches, and machine learning methods. This paper focuses on a variety of optimization techniques to reduce power consumption and resource use

Keywords: Containers, scheduling, optimization and energy consumption

BACKGROUND AND IMPORTANCE OF THE STUDY

INTRODUCTION

In recent years, cloud computing has been widely accepted as a technology that provides a wide range of computer services to individuals, businesses and society via the Internet. The recent expansion of various complex cloud workloads, including cloud memory, machine learning applications, streaming music, secondary services, and more, has led to a significant increase in the need for a variety of cloud services. With Microd services becoming increasingly popular, we expect to see an expansion in cloud services. Virtualization technology is considered an important part of cloud computing, which allows applications to be isolated from the support framework by using the joint use of network, memory and CPU resources to run some programs. From a cloud service provider perspective, cluster management is handled at an abstraction level. This was created by running a container app. Containers hold software with terms, libraries, code and everything you need to do. It's easy to install on any computer, whether it's a real or virtual machine. Container technology uses virtualization in operating systems to move software more flexible and easier. A standard virtual machine (VM) requires all the operating system resources yourself, but the container can use the same operating system core. Tools like Docker Swarm and Kubernetes automate container orchestration across a cluster consisting of management and worker nodes.

Worker nodes run the applications submitted by users, while the management node oversees the entire cluster and the deployment of containers on the worker nodes. The management node also regularly checks the health of the worker nodes to ensure everything operates smoothly. The scheduler plays a key role in distributing workloads efficiently, optimizing performance and resource use.

In cloud data centres, container scheduling involves mapping tasks (T) to containers (C) on physical (PM) or virtual machines (VM) using strategies like bin packing, spreading, and heuristics-based methods. While manual allocation suffices for small-scale setups, automation is essential for scalability. Research focuses on optimizing scheduling algorithms to improve efficiency, performance, and resource utilization in large-scale cloud environments.

This study checks current methods for container planning and highlights gaps and trends in the research. We classify four types of planning algorithms: mathematical approach, heuristic approach, metaheuristic approach, machine learning approach, and analysis of performance metrics in a cloud environment. This study also examines the strengths and limitations of each approach.

II. Containerization

Containers are a new virtualization technique. These allow applications running on the same operating system to be separated into smaller, separate units. A container is a fully packaged portable computer environment that includes everything that your application needs to run. B. Binary files, libraries, configuration files, dependencies. Containers simplify applications by summarizing the complexity of operating systems and resource management. Containerized applications are highly versatile and can function on a variety of infrastructures, including Bar-Metal servers, virtual machines, and cloud platforms.

A. Architecture of Containe :

Containers and virtual machines (VMs) may be similar particularly because they are built on virtualisation technology. But the two are not exactly the same. The main distinction is that containers virtualise operating systems, not the underlying hardware, whereas VMs virtualise physical hardware via a hypervisor.

- The infrastructure consists of hardware components such as CPU, disk storage, and network interfaces.

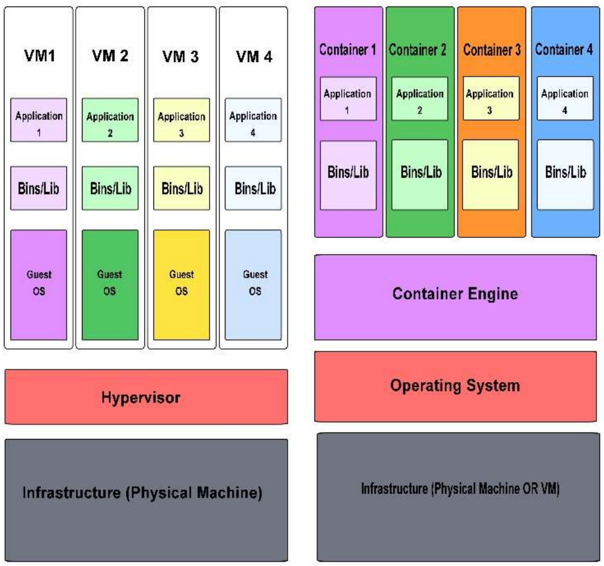

Fig 1 : Virtual Machine Vs Containers

- The bridge between software and hardware is a bridging between hardware and hardware, and the way of life is a bridging between hardware and software.
- Container engines and guest operating systems are specific to containerization technology and run on the host operating system.
- Binary files and libraries are at the top of the list.

When launching an operating system (OS), container-based clouds offer users access to a wider range of applications compared to virtual machine-based clouds. Additionally, they enable application providers to effortlessly handle cloud resources using auto-scaling and migration techniques.

III. SCHEDULING PERFORMANCE METRICS

A. Energy:

Energy consumption is the measure of power used during the deployment of containers on worker nodes. This goal aims to identify schedules that reduce the total power consumption of the cluster. Decreasing power consumption helps cut down operational costs of data centres by lowering energy usage and cooling expenses, leading to higher revenues.

B. Cost:

The total cost of starting the application is set for more services such as storage and processing. Processing costs is time to start the application on available seeds for cluster. The longer you apply for CPUs, the higher the costs.

C. Availability:

It refers to the duration of time when the application is available for use by the user. Access to applications or
services at any time is a crucial aspect of cloud computing. This necessitates a comprehensive planning strategy that incorporates error correction mechanisms for various services in the application, including the provision of redundant containers.

D. Using resources:

This metric measures the efficiency in which sources available for staff nodes such as cores, memory and network ranges are used. In these statistics, the use of resources of the employees of the employees is energy efficient and cheap for the node.

E. Load Compensation:

Workloads from an arithmetic resource perspective should be split evenly and prevent single knots from being overloaded while the other knots remain idle. This balance has implications from response time, cost, and PUT. This is primarily used in microservice architectures, and workloads are often changed.

F. Network Bandwidth:

Defines this number of bits per second that are communicated over the network. Container is used to find network delays and traffic jams during communication.

G. Throughput:

It provides an overview of your system performance (processor, memory, network), but does not guarantee delays or cost reductions.

H. CO2 Footprint:

The carbon dioxide (CO_2) footprint refers to the quantity of carbon dioxide that has been reduced from entering the atmosphere, especially due to the burning of fossil fuels. This figure holds significant importance for the preservation of our environment. Consequently, reducing energy consumption and embracing sustainable energy sources are crucial for minimizing trace CO_2 emissions and decreasing reliance on fossil fuels.

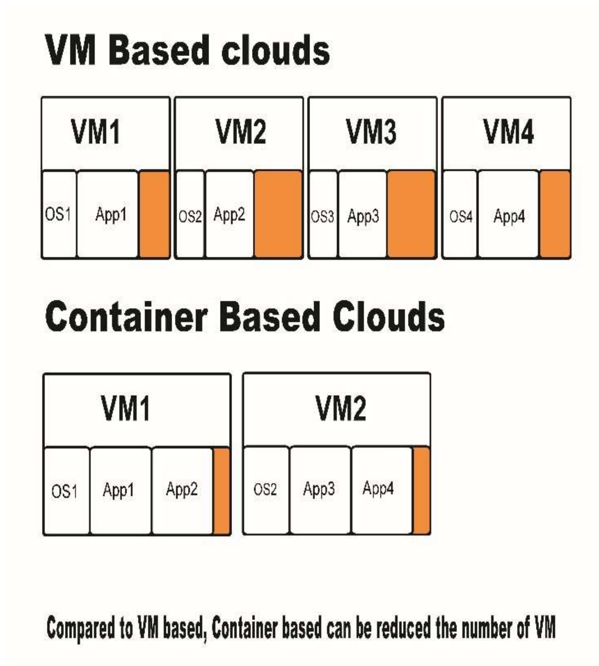

Fig 2 : Resource allocation - Virtual Machine Vs Containers

The primary problem in container scheduling is allocating resources and energy consumption. This can be achieved by allocating resources to the application while ensuring that the load constants of the physical machine remain unchanged. Resource allocation aims to minimize the overall energy usage when distributing resources and decrease the burden on physical machines.

The energy usage of a container cloud data centre is mainly influenced by physical machines, network equipment, cooling systems, and other components. The majority of the energy consumption in a data centre is attributed to physical machines. Consequently, cloud providers aim to decrease the energy usage of their data centres by minimizing the energy consumption of individual physical machines. There are two primary methods to decrease the energy usage of physical mechanical energy.

The first approach is the second goal of reducing the energy consumption of the physical machines themselves, focusing on reducing the number of active physical machines through the movement of the container. In this calculation, PIDLE represents the energy consumed by PM in IDLE, and PMAX refers to energy consumption when fully used.

Energy calculation is done by

TE = Summation of PM energy at every time

Physical machine energy =Pidle+(Pmax−Pidle)×CPU Utilization

IV. Related Works

Integer linear programming (ILP) is a mathematical technique employed to find the best solution for linear functions that are influenced by a set of linear constraints. The main ways to solve ILP problems are two technologies. However, consolidation of container planning problems cannot be resolved in the ppolynomial timeoptimization search space. As a result, heuristic algorithms are often preferred over ILP to obtain solutions more efficiently. Heuristic techniques are used to solve approaches that use practical methods. These methods are scalable and very quick container planning, but do not guarantee the optimality of the solution, but provide near-optimality of the solution. Metaheuristic algorithms have recently taken over leadership in solving difficult optimization problems in a variety of areas. Metaheuristic algorithms are in two categories, including Ant colony optimization (ACO), particle swarm optimization (PSO), whale optimization, and evolutionary algorithms (EAS), among other methods, are commonly used in optimization problems

.

In 2017 [1], the initial linear programming model introduced for the allocation of containers that minimize energy and network expenses. This model showed a cost savings of about. 45% compared to Docker Swarm's binpack strategy. Then, in 2020, [2] introduced a multi-lens image processing model that reduces energy consumption, optimizing its performance. interference and CO2 footprint of the Internet in edge computing environments. This approach went beyond Docker Swarm and Kubernetes, but overlooked container dependencies. In 2021 [4], The problem of efficiently positioning virtual machines (VMP) was tackled to minimize energy usage and prevent resource wastage in cloud computing facilities. They implemented a heuristic algorithm called MINPR. The goal is to decrease power usage by minimizing the number of active physical machines (PMSs) and reducing resource waste. Simulation outcomes demonstrate that MINPR outperforms current techniques. This means that the total energy consumption of a user-defined VM is 10% x 10% of an Amazon EC2 instance. Energy-efficient algorithms focused on the reduction in overload and load expansion objectives. Learning from workload patterns increases energy efficiency compared to cutting-edge methods. In 2020, MoGas proposed a genetic algorithm scheduler with many lenses that optimized container planning for availability, scalability and power consumption. Beyond ACO-based schedulers that improve task distribution by 50%, the clear assignment of tasks is classified by 40% and power consumption is 7%. In 2024, MOMA, an expanded Bi-Lens model of NSGA-II, presented resource use and minimized network efforts. Tested for real Microdo for improved efficiency and reliability.

In 2019 [9] proposed a method to optimize Microdienste failure rates, network transmission and resource utilization using a multi-purpose anis colony plan. It outperforms Spread approach and Genetic algorithm-based approaches in resource balance and reliability but does not consider energy consumption.

In 2023, [10] introduced PS-GWCA, a hybrid PSO-GWO algorithm for microservice container scheduling. GWO avoids local optima, while PSO refines results. PS-GWCA improves network cost (18.07%), load balancing (14.67% local, 20.66% global), search speed (7.5%), and service reliability (5.69%) over existing methods.

In 2019[11], proposed a hybrid model of whale optimization algorithm for optimal container resource allocation, improving utilization, cost efficiency, and system performance over existing methods. In 2023 [12], optimizes container and VM placement in CaaS using the Whale Optimization Algorithm to maximize resource use and minimize power consumption. It efficiently hosts 2000 containers on 633 VMs and 45 PMs, reducing infrastructure needs compared to the best comparison algorithm.

In 2021 [13], the discrete whale optimization (DWO) algorithm was developed to enhance resource allocation and reduce power usage in CaaS (cloud-based automation and storage) environments. Derived from the whale optimization algorithm (WOA), DWO is specifically designed to optimize the placement of containers and virtual machines within cloud data centers.

V. Analysis

The ILP-based techniques discussed in the above, lack scalability and fail to consider all optimization objectives. As problem size grows, solving time increases, limiting their use to small-scale cases. Heuristics are preferred for larger problems, but ILP remains valuable for evaluating heuristic solutions. Heuristics algorithms are also concerned with the optimisation of energy utilisation and resource availability.It is

ABLE 1. COMPARISION OF VARIOUS ALGORITHMS

Reference	**Objectives**				**Methodology**	**Container Technology**	**Strength**
	Energy	***Resource Utilization***	***Network Cost***	***Load balancing***			
[1]	✓		✓		Linear Programming model	Docker	Reduced cost 45% compare with binpack strategy
[2]	✓				Multi-objective ILP-based model	-	Improved latency, energy efficiency, and network performance compared to traditional scheduling methods.
[3]		✓	✓		Heuristic	-	Min-cost flow mode-Heterogeneous resources
[4]	✓	✓			Heuristic	C++ - Small Setup heterogeneous clouds	Minimizing power consumption and optimizing resource utilization through a

							multi-dimensional resource usage model.
[5]	✓			✓	Heuristic	Cloudsim	Minimizing energy computation of host.
[6]	✓	✓			Genetic Algorithm	Java	Online resource allocation methods used to reduce energy and maximize resource utilization.
[7]	✓	✓			Genetic algorithms	Swarm kit – GO programming language	It mainly concentrates on resource availability and cut power use.
[8]		✓			Genetic algorithms	Kubernetes	Resource allocation
[9]		✓	✓	✓	Ant colony optimization	-	Reduce network cost Use resources efficiently. No energy reduction.
[10]			✓	✓	Hybrid algorithm PSO +GWO	Docker	Improves network cost, load balance, search availability
[11]	✓	✓		✓	Hybrid model multi optimization technique (Whales + Lion)	Matlab	**over-provisioning** and **under-provisioning** of resources

							better performance and resource optimization.
[12]	✓	✓		✓	Whales optimization + DCP approaches	Java Cloudsim	It outperforms in energy consumption, resource allocation and cost
[13]	✓		✓	✓	Whales optimization	Cloudsim	Improves energy consumption,loadbalancing and balanced cost.

faster than linear formulation. It also integrates with other optimisation algorithms to increase the quality of solutions. Meta-heuristic algorithms are well-suited for multi-objective optimization. However, when dealing with large-scale problems, their computational complexity necessitates parallelization on platforms such as GPU clusters to reduce execution time.

VI. CONCLUSION

Containers quickly become the foundation of modern cloud computing, enabling the development of applications in a variety of domains, including IoT, microservices, autonomous vehicles, and intelligent infrastructure. An effective container planning is extremely important for optimizing cloud resource utilization over time. However, a detailed analysis of existing strategies in container planning strategies shows that a single algorithm can fully achieve all performance objectives, highlighting the need for further research and innovation in this field.

As a light virtualization technology, containers play an important role in providing services in Cloud and edge computing environments. Given the fact that edge devices often have limited computing power and limited battery life, it is essential to design resource-conscious, energy-efficient container planning methods to ensure optimal performance and sustainability in this rapidly growing application domain.

The multi-purpose metaheuristic algorithm illustrates a promising approach to comprehensive resource management. However, by incorporating local search techniques into sophisticated early generation populations, we can improve solution quality and accelerated convergence. Furthermore, by inserting hybrid forms such as integration of heuristics into metaheuristics and combining two different metaheuristic selves, the search chamber can be examined more efficiently and effectively. This advancement greatly improves the adaptability and robustness of container planning mechanisms, paving the way for more intelligent and autonomous cloud resource

REFERENCES

[1] Dong Zhang, Bing-Heng Yan, Zhen Feng, Chi Zhang, and Yu-Xin Wang. Container oriented job scheduling using linear programming model. In 2017 3rd International Conference on Information Management (ICIM), pages 174–180, 2017.

[2] Kuljeet Kaur, Sahil Garg, Georges Kaddoum, Syed Hassan Ahmed, and Mohammed Atiquzzaman. Keids: Kubernetes-based energy and interference driven scheduler for industrial iot in edge-cloud ecosystem. IEEE Internet of Things Journal, 7(5):4228–4237, 2020.

[3] Hu, Y., Zhou, H., de Laat, C., Zhao, Z., 2020. Concurrent container scheduling on heterogeneous clusters with multi-resource constraints. Fut. Gen. Comput. Syst. 102, 562–573.

[4] Azizi, S., Li, D., et al.: An energy-efficient algorithm for virtual machine placement optimization in cloud data centers. Clust. Comput. 23, 3421–3434 (2020). https://doi.org/10.1007/s10586 020-03096-0

[5] Saadi Y, El Kafhali S (2020) Energy-efficient strategy for virtual machine consolidation in cloud environment. Soft Comput 24:14845–14859

[6] Boxiong Tan, Hui Ma, Yi Mei, and Mengjie Zhang. A cooperative coevolution genetic programming hyper-heuristic approach for on-line resource allocation incontainer-based clouds. IEEE Transactions on Cloud Computing, pages 1–1, 2020.

[7] Mahmoud Imdoukh, Imtiaz Ahmad, and Mohammad Alfailakawi. Optimizing scheduling decisions of container management tool using many objective genetic algorithm. Concurrency and omputation: Practice and Experience, march 2020.

[8] Q. -H. Chen and C. -Y. Wen, "Optimal Resource Allocation Using Genetic Algorithm in Container-Based Heterogeneous Cloud," in *IEEE Access*, vol. 12, pp. 7413-7429, 2024.

[9] Lin, M., Xi, J., Bai, W., Wu, J., 2019. Ant colony algorithm for multi-objective optimization of container-based microservice scheduling in cloud. IEEE Access 7, 83088–83100.

[10] X. Chen, Y. Wu and S. Xiao, "Particle Swarm–Grey Wolf Cooperation Algorithm Based on Microservice Container Scheduling Problem," in *IEEE Access*, vol. 11, pp. 16667-16682, 2023, doi: 10.1109/ACCESS.2023.3244881.

[11] Vhatkar, K.N., Bhole, G.P., 2019. Optimal container resource allocation in cloud architecture: A new hybrid model. Journal of King Saud University-Computer and Information Sciences

[12] Alwabel, A. A Novel Container Placement Mechanism Based on Whale Optimization Algorithm for CaaS Clouds. Electronics 2023, 12, 3369. https://doi.org/10.3390/electronics12153369

[13] Al-Moalmi, A.; Luo, J.; Salah, A.; Li, K.; Yin, L. A whale optimization system for energy-efficient container placement in data centers. Expert Syst. Appl. 2021, 164, 113719