Survey on Energy Consumption Optimization Approach in Container Based Cloud Environments

Francesco Pace Napoleone, Marian-Catalin Mutu-Costan, Ryzhuk Ihor, Ewan Soueidan
Université de Toulouse

Abstract
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords

Cloud Computing · Containerization · Energy Optimization · Microservices · Resource Allocation.

Contents

1 Introduction

Cloud computing evolved to support scalable platform usage, moving from single-file applications to client/server architectures with monolithic backends, then to microservices with containers, and now transitioning to micro-frontends. Containers run directly on the host using Linux CGroups, bypassing the hypervisor overhead inherent in VMs (e.g., Proxmox, VMWare), which allows for full utilization of host CPU resources and simplifies telemetry monitoring. This lightweight nature is evident in Docker files that simply copy files and run builds, making containers easier to integrate into CI/CD pipelines.

However, this increased reliance on cloud-based infrastructures has led to a significant rise in energy consumption. In 2006, the electricity costs for IT infrastructures in the United States alone were estimated at $4.5 billion, with projections to double by 2011[4]. Energy consumption optimization has since become a critical concern, especially as cloud data centers now account for approximately 1-1.5% of global electricity use[11]. Despite efficiency improvements, the demand for digital services continues to grow, pushing the need for more sustainable solutions.

Early research focused on heuristic-based approaches to optimize virtual machine (VM) placement, achieving energy savings of up to 83% while maintaining only a 1.1% service level agreement (SLA) violation rate[4]. More recently, research has shifted from VM-based allocation towards containerized environments, where energy efficiency is influenced by scheduling strategies, workload distribution, and infrastructure optimizations. Studies indicate that modern cloud providers, including Amazon, Microsoft, Google, and Meta, have doubled their energy consumption between 2017 and 2021, reaching 72 TWh in 2021[14;9;11].

The paper "Survey on Energy Consumption Optimization Approach in Container Based Cloud Environments" further highlights that containerization not only drives scalability and reproducibility but also plays a crucial role in optimizing energy consumption. It explores strategies for efficient resource allocation, reducing power overhead, and ensuring that the benefits of container-based deployments extend beyond performance to sustainability in cloud infrastructures.

In this work, we present a state-of-the-art review on Energy Consumption Optimization Approaches in Container-Based Cloud Environments. Our survey of the available literature—predominantly spanning from 2010 to 2020—reveals that foundational research primarily focused on energy measurement, basic optimization strategies, and energy visualization techniques[4].

Early contributions, such as those by Beloglazov and Buyya[4], as well as Piraghaj et al.[15], laid the groundwork for dynamic resource allocation and energy-efficient container consolidation. Later advancements introduced more sophisticated container scheduling mechanisms, including availability-aware scheduling[1], concurrent scheduling in heterogeneous clusters[10], and hybrid AI-driven resource allocation[17].

In parallel, energy-efficient resource management techniques gained prominence, incorporating renewable energy-aware scheduling[12], optimization-based consolidation methods[16;15], and brownout-based scheduling strategies[18]. Several studies further explored predictive optimization and SLA-aware provisioning frameworks to enhance energy efficiency[7;8;13;5;6].

Beyond optimization techniques, researchers have also examined broader energy consumption trends and policy implications[2], reflecting the increasing emphasis on sustainability in cloud computing. Additionally, efforts in DevOps-driven elastic container management have contributed to improving the adaptability and efficiency of containerized cloud applications[3].

Although these prior works have significantly advanced the field, our review highlights an evolving trend toward more integrated and user-centric energy management strategies, as reflected in recent data on energy consumption in cloud environments[14;9;11].

2 Methodology

Systematic Procedure to Identify Papers

The research process started with the defintion of the topic together with our supervisor Professor Jean-Marc Pierson, together we decided to settle on creating a state fo the art review on the topic of Container and Cloud Computing, with a particular emphasis on energy consumption optimization. To begin with we defined a "Search Pipeline"

Figure 1: Search Pipeline Flowchart

This procedures allowed us to organize the work and have a standardized method of acceptance for navigating the papers. We also defined a search string with which we would start the search: ((energy OR resource) AND container). We decided to expand the search also to resource since we noticed that many times better resource utilization leads to better dynamic power handling, hence lowering the amount of energy required to carry out a task. Before starting with the research we also defined some acceptance criteria that would define whether a paper was going to be included in our survey or not. We defined the criteria as follows:

Direct Exclusion of Paper If
Work is not in English
Work is not a scientific paper
Work has less than 10 citations
Work is not about containers/cloud

Table 1: Exclusion Rules

Hence we proceeded with our search and we identified a total of 34 papers. That are divided as follows through the various publishers

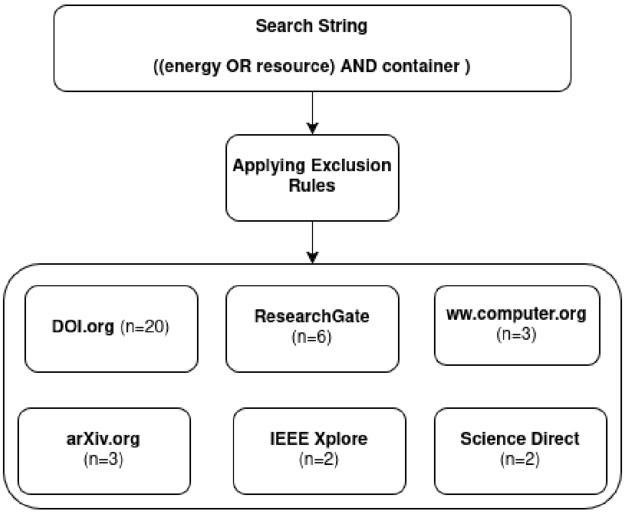

Figure 2: Exported Papers

To have an insight on the publication dates for the filtered papers, we can clearly visualize how the period between 2017 and 2019 was the most important for research in this field.

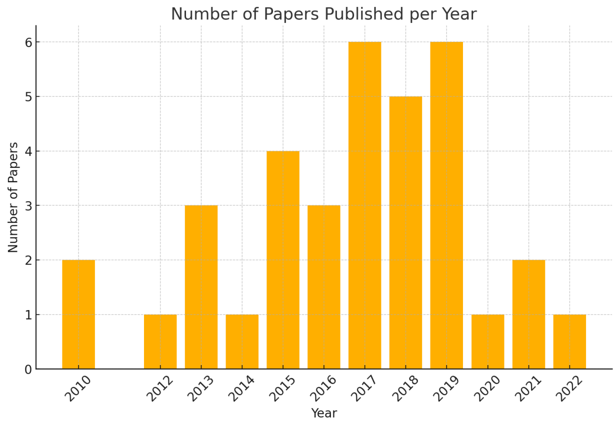

Figure 3: Papers per year

At this point we could move to the classification step of our workflow. We decided to define the following metrics

Classification Categories
Optimization Technique
Target Level
Scheduling Scope
Energy Strategy
Docker Awereness

Table 2: Paper Classification

The selected classification metrics capture the core technical dimensions of energy-aware scheduling research. Optimization Technique highlights the algorithmic strategy used. Target Level distinguishes whether the solution operates at the container, VM, or physical machine level. Scheduling Scope captures whether the approach is reactive, predictive, or concurrent. Energy Strategy defines the method used to reduce energy consumption, such as consolidation or renewable-aware placement. Finally, Docker Awareness situates each work in the context of containerization maturity, relevant for comparing approaches across technological shifts.

Paper (Short Title)	Optimization Algorithm	Target Level	Scheduling Scope	Energy Strategy	Docker Awareness
Availability-Aware Scheduler [1] (2018)	Greedy	Container, VM	Offline	Availability	Mature Docker
Renewable-Aware Scheduling [12] (2019)	Greedy	Container, VM	Reactive	Consolidation	Mature Docker
Concurrent Container Scheduling [10] (2020)	MCMF	Container	Concurrent	Consolidation	Mature Docker
GP Hyper-Heuristic for Containers [17] (2019)	Evolutionary	Container, VM	Online	Consolidation	Mature Docker
PSO-Based Container Consolidation [16] (2018)	PSO	Container	Offline	Consolidation	Mature Docker
Energy-Efficient Container Framework [15] (2015)	Greedy	Container, VM	Reactive	Consolidation	Early Docker
VM Consolidation Strategy [6] (2017)	Meta-Heuristic	VM	Online	Consolidation	Mature Docker
Predictive Resource Provisioning [7] (2015)	Model-based	VM	Predictive	Consolidation	Early-Docker
Dynamic VM Reallocation [4] (2010)	Greedy	VM	Online	Consolidation	Pre-Docker
Autonomic Container Management [3] (2017)	Model-based	Container, VM	Predictive	QoS	Mature Docker
Brownout-Based Scheduling [18] (2016)	Greedy	Application Layer	Reactive	Brownout	Mature Docker
GPR + Convex VM Planning [5] (2017)	Model-Based	VM, PM	Predictive	Consolidation	Mature Docker
SLA-Aware Consolidation [13] (2018)	Model-Based	VM, PM	Predictive	Consolidation	Mature Docker

Table 3: Classification of Technical Papers on Energy-Aware Scheduling

Paper Ref.	Type	Scope / Focus
[8] Survey on Energy-Aware Scheduling	Survey	Overview of VM and container-level techniques
[14] Global Energy Estimates	Analysis	Environmental impact of data centers
[9] Cloud Impact on Energy	Report	Energy use in European data centers
[11] IEA Report on Data Centers	Report	Global energy trends in ICT and data centers
[2] EU Data Center Energy Trends	Analysis	EU-level trends in ICT energy consumption

Table 4: Survey and Analysis Papers

3 Trend reconstruction

Chronological evolution of the trend

3.1 Early approaches

Early approaches date back to 2010, with Docker being released in 2013. In 2006, the cost of electricity consumed by IT infrastructures in the US was estimated at 4.5 billion dollars and was projected to double by 2011 [4].

In the initial attempts described in [4], a simple Bin Packing variation with DVFS enabled was used to address this problem. The results, obtained using the CloudSim environment, showed an energy savings gain of 83% compared to no policy, with an SLA violation of 1.1%. This marked the beginning of research in this area.

In 2015, researchers shifted toward predictive modeling to address the limitations of static thresholds. Dabbagh et al. [7] proposed a Wiener filter-based predictor to estimate cluster workloads, combined with Best Fit Decreasing for PM allocation. Their approach improved energy efficiency by up to 33% compared to heuristic-based methods.

Then, in 2017, further improvements were made to predictive methods, demonstrating the rapid progress in this field. Bui et al. [5] advanced previous work by employing Gaussian Process Regression (GPR) for non-stationary workload prediction, accelerated using Fast Fourier Transform to address the computational complexity of GPR. Additionally, they integrated convex optimization-based migration. Their model achieved an even higher energy efficiency improvement—35% compared to heuristic approaches—while maintaining low latency, with only a 15% latency trade-off.

3.2 Containers Rise

Containers began replacing VMs for lightweight virtualization. Initial efforts focused on container migration to reduce active VM count. Strategies involved modular watchdogs and Pearson correlation checks. [Notes - Long Version.pdf]

Flow network models enabled scalable concurrent container scheduling with multi-resource awareness and affinity considerations. These models balanced execution speed and resource utilization across up to 5000 machines. [Notes - Long Version.pdf]

Focus shifted to application-layer availability using metrics like MTTF and MTTR. Scheduling considered redundancy, affinity, and anti-affinity rules to ensure SLA compliance during failures. [Notes - Long Version.pdf]

---

4 Method Analysis

Conclusion

4.1 Allocation

HERE WE SHOULD HAVE TABLES THAT SHOW THE NUMBERS FOR ALL METHODS USED. WHICH ALGORITHMS OVER THE YEARS, THE COMPLEXITY, PARAMETERS TAKEN INTO ACCOUNT

4.2 Evaluation

HERE WE SHOULD HAVE TABLES/descriptions comparing different sizes of the evaluation. Meaning that we should compare the hardware and environment each approach was tested on. IF WE DO THIS WELL, WE HAVE THE RESULTS/CONCLUSION SECTION FOR FREE

---

5 Conclusions

HERE WE SHOULD COMPARE THE DIFFERENT APPROACHES AND CHOOSE WINNERS(?)

References

[1] Yanal Alahmad, Tariq Daradkeh, and Anjali Agarwal. *Availability-Aware Container Scheduler for Application Services in Cloud.*

[2] Maria Avgerinou, Paolo Bertoldi, and Luca Castellazzi. Trends in data centre energy consumption under the european code of conduct for data centre energy efficiency. 10(10):1470.

[3] Cornel Barna, Hamzeh Khazaei, Marios Fokaefs, and Marin Litoiu. *Delivering Elastic Containerized Cloud Applications to Enable DevOps.*

[4] Anton Beloglazov and Rajkumar Buyya. Energy efficient allocation of virtual machines in cloud data centers. In *2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, pages 577–578. IEEE.

[5] Dinh-Mao Bui, YongIk Yoon, Eui-Nam Huh, SungIk Jun, and Sungyoung Lee. Energy efficiency for cloud computing system based on predictive optimization. 102:103–114.

[6] Alessandro Carrega and Matteo Repetto. Energy-aware consolidation scheme for data center cloud applications. In *2017 29th International Teletraffic Congress (ITC 29)*, pages 24–29. IEEE.

[7] Mehiar Dabbagh, Bechir Hamdaoui, Mohsen Guizani, and Ammar Rayes. Energy-efficient resource allocation and provisioning framework for cloud data centers. 12(3):377–391.

[8] Abdul Hameed, Alireza Khoshkbarforoushha, Rajiv Ranjan, Prem Prakash Jayaraman, Joanna Kolodziej, Pavan Balaji, Sherali Zeadally, Qutaibah Marwan Malluhi, Nikos Tziritas, Abhinav Vishnu, Samee U. Khan, and Albert Zomaya. A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems. 98(7):751–774.

[9] Ralph Hintemann and Simon Hinterholzer. Cloud computing drives the growth of the data center industry and its energy consumption. *Data centers*, 2022.

[10] Yang Hu, Huan Zhou, Cees de Laat, and Zhiming Zhao. Concurrent container scheduling on heterogeneous clusters with multi-resource constraints. 102:562–573.

[11] IEA. Data centres & networks. `https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks`. Accessed: 2025-03-25.

[12] Neeraj Kumar, Gagangeet Singh Aujla, Sahil Garg, Kuljeet Kaur, Rajiv Ranjan, and Saurabh Kumar Garg. Renewable energy-based multi-indexed job classification and container management scheme. 15(5):2947–2957.

[13] Lianpeng Li, Jian Dong, Decheng Zuo, and JIaxi Liu. SLA-aware and energy-efficient VM consolidation in cloud data centers using host states naive bayesian prediction model. pages 80–87. IEEE Computer Society.

[14] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. Recalibrating global data center energy-use estimates. *Science*, 367(6481):984–986, 2020.

[15] Sareh Fotuhi Piraghaj, Amir Vahid Dastjerdi, Rodrigo N. Calheiros, and Rajkumar Buyya. A framework and algorithm for energy efficient container consolidation in cloud data centers. In *2015 IEEE International Conference on Data Science and Data Intensive Systems*, pages 368–375. IEEE.

[16] Tao Shi, Hui Ma, and Gang Chen. *Energy-Aware Container Consolidation Based on PSO in Cloud Data Centers.*

[17] Boxiong Tan, Hui Ma, and Yi Mei. *A Hybrid Genetic Programming Hyper-Heuristic Approach for Online Two-level Resource Allocation in Container-based Clouds.*

[18] Minxian Xu, Amir Vahid Dastjerdi, and Rajkumar Buyya. Energy efficient scheduling of cloud application components with brownout. 1(2):40–53.