OPTIMIZING VIRTUAL MACHINE PLACEMENT IN CLOUD DATA CENTERS: ENHANCED ANT COLONY OPTIMIZATION APPROACH

Kelvin Ovabor and Travis Atkison

The University of Alabama, Tuscaloosa, Alabama, USA

ABSTRACT

In cloud computing, efficient resource allocation within data centers is crucial for reducing energy consumption and operational costs. Virtual Machine Placement (VMP) isa critical aspect, involving the strategic assignment of Virtual Machines (VMs) to physical servers. However, inefficient VMplacement can lead to increased energy usage, posing significant challenges to operational efficiency and cost-effectiveness. This paper introduces a novel approach to VM placement, with the aim of minimizing total energy consumption within data centers. Leveraging the Ant Colony Optimization (ACO)algorithm, we customized its information heuristic based on the energy efficiency of physical machines (PMs) within data centers. Experimental validation demonstrates the scalability ofour approach in large data center environments, where it notably out performs the selected benchmark, the ACOVMP (Ant Colony Optimization Virtual Machine Placement) algorithm, in terms of energy consumption. Our findings highlight the effectiveness ofour approach in optimizing VM placement decisions, contributing to ongoing efforts to enhance energy efficiency and operational sustainability in cloud data center environments.

KEYWORDS

Cloud, Virtual Machine, Ant Colony Optimization, Data Center, Energy Consumption

1. INTRODUCTION

Cloud computing has revolutionized the landscape of modern IT infrastructure, offering scalable and flexible computing resources to meet the ever-growing demands of digital services and applications [1], [2]. Central to this paradigm shift are cloud data centers, which form the backbone of the cloud computing ecosystem, tasked with efficiently managing and allocating resources to ensure optimal performance and cost-effectiveness [3], [4]. However, the rapid proliferation of cloud data centers has introduced significant challenges, foremost among them being efficient resource utilization and the mitigation of energy consumption and operational costs [5], [6].

In this context, Virtual Machine Placement (VMP) emerges as a critical aspect of resource management, involving the strategic allocation of Virtual Machines (VMs) to physical servers within data centers [7], [8]. Optimizing VM placement is essential not only for maximizing resource utilization but also for minimizing energy consumption and operational expenses. Inefficient VM placement can lead to sub-optimal resource allocation, resulting in increased energy consumption and reduced operational efficiency [9], [10]. Moreover, as data centers

David C. Wyld et al. (Eds): Signal, CDKP, AdNLP, MLT, SAI – 2025
pp. 159-168, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.150912

continue to scale in size and complexity, the need for sophisticated optimization techniques to address the VMP problem becomes increasingly apparent [11].

Traditional methods for tackling the VMP problem, such as linear programming, often struggle to cope with the scale and combinatorial intricacies of modern cloud environments. Consequently, heuristic and metaheuristic approaches have gained traction due to their ability to efficiently explore large solution spaces [12], [13]. Among these, Ant Colony Optimization (ACO) stands out as a promising metaheuristic inspired by the foraging behavior of ants, known for its effectiveness in solving NP-hard optimization problems, including VMP [14], [15].

In this paper, we address the challenge of inefficient VM placement within cloud data centers and its implications on energy consumption and operational costs. Building upon existing research, we propose a novel ACO-based approach to VM placement aimed at minimizing total energy consumption within data centers. Our approach is enhanced by tailored heuristics that prioritize physical machines (PMs) based on their energy efficiency profiles. Through extensive experimental validation in large-scale data center scenarios, we demonstrate the efficacy of our method. Comparative analyses against benchmark algorithms highlight its scalability and superior performance in optimizing VM placement decisions, thereby contributing to the broader goal of improving energy efficiency and operational effectiveness in cloud data centers.

2. LITERATURE REVIEW

Efficient resource management within cloud data centers is of paramount importance, as it directly influences energy consumption and operational expenses. A significant aspect of this endeavor is Virtual Machine Placement (VMP), where the allocation of virtual machines (VMs) to physical servers is meticulously orchestrated to ensure seamless operations [1], [3], [4]. However, VMP presents a formidable challenge akin to solving a complex puzzle, with conventional methods often struggling to yield optimal solutions [2], [5].Ant Colony Optimization (ACO) emerges as a sophisticated approach for addressing such intricate problems, drawing inspiration from the efficient foraging behavior observed in ants. By harnessing ACO, computer algorithms can emulate the communication and problem-solving prowess of ants, offering a potent tool for tackling tasks like VM placement with precision [6], [7]. Previous research endeavors have explored the application of ACO in VM placement, demonstrating promising results. For instance, Liu et al. (2017) devised a method leveraging ACO to enhance VM placement intelligence, considering factors such as resource requirements, server capacities, and network efficiency [9]. Similarly, Zhang and colleagues (2019) augmented ACO with simulated annealing, elevating the efficacy of VM placement algorithms to handle complex scenarios efficiently [10]. More recently, Wang et al. (2023) proposed an energy-aware ACO strategy that further improves VM placement with a focus on reducing energy consumption in cloud data centers [8].Despite its efficacy, challenges persist in optimizing ACO for dynamic environments characterized by rapid changes in workload demand. Addressing these challenges necessitates refining ACO algorithms to adapt swiftly to fluctuations in demand, alongside integrating considerations for energy conservation and environmental sustainability into decision-making processes [3], [4], [5], [11]. Contemporary studies have also begun exploring the integration of machine learning and reinforcement learning techniques with ACO to enhance adaptability and efficiency in resource scheduling [6], [7].

In this study, we leverage insights gleaned from prior research to advance the application of ACO in VM placement within data centers. Our approach aims to optimize VM placement decisions to minimize energy consumption while enhancing operational efficiency. We augment ACO with tailored strategies to prioritize servers with lower energy utilization, thereby contributing to the overarching objective of making data centers more energy-efficient and cost-effective. Through

rigorous testing in large-scale data center environments, we seek to validate the efficacy of our approach and its potential to drive tangible improvements in data center operations.

3. ENERGY CONSUMPTION IN CLOUD DATA CENTERS

Energy consumption in cloud data centers is a multiface issue influenced by several critical components, including server operations, cooling systems, networking equipment, and all other facility infrastructure. Among these, servers remain the primary consumers of energy, accounting for the largest share of total power usage within data centers [1], [2]. The continuous operation of servers to handle computational tasks such as processing user requests, running applications, and managing data storage leads to significant energy demands. Cooling systems are equally essential, as they maintain optimal environmental conditions by dissipating the heat generated by servers and other hardware. These cooling infrastructures consume substantial energy to regulate temperature and humidity, thereby ensuring equipment reliability and longevity [3], [4]. Additionally, networking equipment—including switches, routers, and cabling contributes to the energy footprint by enabling data transmission and connectivity within and beyond the data center environment [5].

Statistical data and industry trends further highlight the scale of this energy consumption challenge. For instance, global data center electricity consumption was estimated at approximately 201.8 terawatt-hours (TWh) in 2010, with projections indicating a steady and significant increase driven by the exponential growth of cloud services, data storage demands, and computational resource utilization [6], [7]. Recent studies confirm this upward trajectory, emphasizing the urgent need for energy-efficient resource management strategies and advanced optimization techniques to mitigate the environmental and operational costs associated with data center energy consumption [8], [9]. Figure 1 illustrates this growing trend, underscoring the importance of innovative approaches such as energy-aware virtual machine placement and infrastructure management in addressing this critical issue.

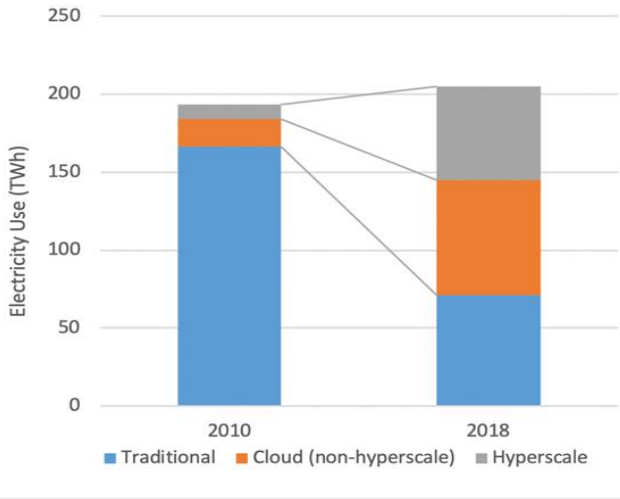

Fig. 1. Estimated Global Data Electricity Used By Data Centers 2010 and 2018. Source: Masanet *et al.* 2020.

Furthermore, studies have shown that server utilization rates in cloud data centers typically range from 11% to 50%,indicating inefficiencies in resource utilization and energy consumption [5]. Idle servers consume a substantial amount of power, with estimates suggesting that an active but

idle server may consume between 50% and 70% of the power consumed by a fully utilized server [6]. These inefficiencies underscore the importance of implementing energy-efficient practices and optimization strategies in cloud data centers to mitigate energy consumption and reduce operational costs.

4. APPLICATION OF NATURE-INSPIRED ALGORITHMS

Nature-inspired algorithms offer versatile and efficient solutions for optimizing various aspects of cloud data centero perations, ultimately reducing energy consumption. These algorithms draw inspiration from natural phenomena and biological systems to devise innovative optimization techniques. In this section, we explore how nature-inspired algorithms can be applied to optimize different aspects of cloud data center operations:

4.1. Workload Scheduling

Nature-inspired algorithms, such as Genetic Algorithms(GA) and Particle Swarm Optimization (PSO), can optimize workload scheduling in cloud data centers. These algorithms mimic the evolutionary process or swarm behavior observed in nature to iteratively search for optimal task allocation strategies. By considering factors like task dependencies, resource availability, and energy efficiency objectives, nature-inspired algorithms can dynamically schedule workloads across servers to minimize energy consumption while meeting performance requirements.

4.2. Resource Allocation

Nature-inspired algorithms can optimize resource allocation in cloud data centers by dynamically assigning virtual machines (VMs) to physical servers. Algorithms like Ant Colony Optimization (ACO) and Simulated Annealing (SA) emulate the foraging behavior of ants or the annealing process observed in metallurgy to explore and optimize resource allocation configurations. By considering factors like server utilization, workload characteristics, and energy efficiency goals, these algorithms can efficiently distribute computational tasks to minimize energy consumption and improve resource utilization.

4.3. Cooling System Optimization

Nature-inspired algorithms can optimize cooling system operation in data centers to maintain optimal temperature and humidity levels while minimizing energy consumption. Algorithms like Genetic Algorithms (GA) and Artificial BeeColony (ABC) optimization mimic natural selection or swarm behavior to optimize cooling system parameters such as fan speed, airflow direction, and temperature set points. By dynamically adjusting cooling system settings based on realtime data and environmental conditions, these algorithms can reduce energy consumption associated with data center cooling while ensuring equipment reliability and performance.

4.4. Hardware Management

Nature-inspired algorithms offer powerful solutions for optimizing hardware management in cloud data centers by dynamically adapting hardware configurations to fluctuating workload demands and energy efficiency targets. Techniques such as Genetic Algorithms (GA) and Evolutionary Strategies (ES) mimic the natural evolutionary process, iteratively evolving optimal hardware setups based on key performance indicators and energy consumption goals. These algorithms effectively consider critical factors including hardware heterogeneity, varying

workload characteristics, and power usage patterns, enabling intelligent optimization of hardware provisioning and utilization.

By harnessing the adaptive and self-organizing principles observed in natural and biological systems, these algorithms facilitate improvements across several key domains: workload scheduling, resource allocation, cooling system operation, and overall hardware management. This integrated optimization approach helps reduce energy consumption while boosting system performance and sustainability within cloud computing infrastructures. Recent studies have demonstrated that leveraging such nature-inspired meta heuristics can achieve significant energy savings and contribute to the development of greener, more sustainable cloud data center operations [11], [12].

5. FORMULATION OF THE VM PLACEMENT PROBLEM

A cloud data center comprises numerous physical machines (PMs), each varying in CPU and memory capacities, as well asenergy efficiency. Multiple virtual machines (VMs) need to be deployed on these PMs, each with its own CPU and memory requirements, along with different arrival and execution times. The proposed approach involves allocating these VMs tohosting PMs over various time intervals while ensuring that resource capacities are satisfied. Below is the formulation forVM placement

A. Inputs

The algorithm is fed the following inputs obtained from the system profile:

• Physical Machines (PMs):

– PMs are denoted as pmj.
– Each PM has its capacity defined by CPU (CPU_pmj), memory (RAM_pmj), and maximum energy consumption (pm.Max).

• Virtual Machines (VMs):

– VMs are denoted as vmi.
– Each VM has requirements for CPU (CPU_vmi) and memory (RAM_vmi), as well as arrival time and execution time.

• Binary Variable Placement of VMs on PMs is represented by a binary variable Xij, where i is the VM index and j is the PM index. This variable indicates whether VM i is placed on PM j or not:

Xij =
(
1 if vmi is allocated to pmj
0 if vmi is not allocated to pmj
)

B. Objective

The objective is to minimize the total energy consumption of the data center operating for 24 hours.

C. Constraints

1) CPU Capacity Constraint: Ensures the total CPU usage of the VMs on a PM does not exceed its CPU capacity:
$\sum i\ Xij \times CPU\_vmi \leq CPU\_pmj\ \forall\ pmj$

2) RAM Capacity Constraint: Similar to CPU, ensures RAM requirements of VMs do not exceed the PM's capacity:
$\sum i\ Xij \times RAM\_vmi \leq RAM\_pmj\ \forall\ pmj$

3) VM Allocation Constraint: Ensures each VM is allocated to exactly one PM:
$\sum j\ Xij = 1\ \forall\ vmi$

These constraints ensure that the allocation of VMs to PMs is feasible and satisfies the resource requirements of both VMs and PMs.

D. Energy Cost Calculation

The energy consumption of physical machines (PMs) has a direct linear relationship with CPU utilization. The formula used is:

Energy cost = $\sum j$ (CPU usage_pmj × Energy unit) × T
E. CPU Utilization
Due to the variability of workload on VMs, each time interval is divided into small slots. The CPU utilization is calculated as:
CPU utilization = $\sum i$ CPU usage_vmi / T

This section describes the formulation of the VM placement problem, including inputs, constraints, objective function, and energy cost calculation in cloud data centers.

6. EXPERIMENTAL DESIGN AND RESULT EVALUATION

The simulations were carried out using the ACO simulator tailored to the proposed algorithm. This simulator provides a platform for implementing and testing Ant Colony Optimization (ACO) algorithms in various optimization scenarios, including virtual machine placement in cloud data centers. The simulator allows researchers to configure parameters specific to their algorithm, such as pheromone update rules, ant behavior, and problem representation. In our case, we customized the ACO simulator to implement the Proposed Ant Colony Optimization for Virtual Machine Placement (PAVM) algorithm. This involved configuring parameters such as the number of ants, the pheromone evaporation rate, and the heuristic information used by ants to make decisions about virtual machine placement. Additionally, we adapted the simulator to handle the heterogeneous environment of the data center, where physical machines have varying capacities and energy efficiencies. By utilizing the ACO simulator, we were able to conduct rigorous experiments to evaluate the performance of the PAVM algorithm under different data center configurations and workload scenarios. Specifically, we conducted five experiments, incrementally increasing the number of VMs from 2 to 5 in steps of 1,and simultaneously increasing the number of PMs from 500 to 3000 in steps of 500. These experiments provided valuable data for assessing the algorithm’s scalability, efficiency, and ability to minimize energy consumption while optimizing virtual machine placement. The implementation of simulation for both the proposed approach PAVM and the benchmark algorithm AVOCMP was coded using Python on a desktop computer running Windows 10. The computer was equipped with an Intel Core i7-4790 CPU (3.60 GHz) and 16 GB RAM. Table I outlines the specific test problems

employed in these experiments. The use of Python facilitated the experimentation process, enabling efficient customization and analysis of the results.

PAVM (Proposed Ant Colony Virtual Machine) Algorithm for VM Placement
Algorithm: PAVM for VM Placement
Input: Set of PMs, set of VMs V, set of ants antSet, Set of parameters
Output: Best solution of VM placement

1) for nAnt = 1 to Ant number do
2) Initialize data structure for each ant;
3) for interval = 1 to total number of intervals do
4) Sort VMs list in each interval in descending order of CPU requirements;
5) Shuffle VMs list;
6) //Algorithm starts
7) Energytotal ← 0.0;
8) for interval = 1 to total number of intervals do
9) Energyinterval ← 0.0;
10) Sort active PMs in the interval in descending order based on Energy Efficiency;
11) Sort inactive PMs in the interval in descending order based on Energy Efficiency;
12) Add active and inactive PMs to PMs list;
13) Initialize $\tau 0$ = FFD Solution for this Interval using Equation (5);
14) // ACO Starts
15) Initialize ACO parameters: nCycleNoImp = 0, ant number;
16) for iteration until no improvement do
17) for iAnt = 0 to ant number do
18) Initialize VMList, PMlist, $\tau i,j$ for the current ant;
19) // compute solution
20) for i = 0 to VM number in the current interval do
21) for j = 0 to PM number do
22) if i VM is feasible in PM j then
23) Calculate information heuristic using Equation (6);
24) Calculate pseudo-random-proportional using Equation (7);
25) Calculate random-proportional rule using Equation (8);
26) Generate a random number q;
27) if q < q0 then
28) Choose a PM using Equation (7);
29) else
30) Choose a PM using Equation (8);
31) Calculate the solution;
32) if the solution improved then
33) bestsolution = currentsolution;
34) Reset nCycle = 0;
35) else if nCycle > iterationMaxCondition then
36) break;
37) else
38) nCycle = nCycle + 1;
39) // global Pheromone Update
40) for i = 0 to VM number do
41) for i = 0 to PM number do
42) update solutions using Equation (9);
43) Output the best solution for the interval;
44) Energyinterval = Energy result from interval solution;

45) Energytotal += Energyinterval;
46) Output the final solution for each interval.
47.Output: Best solution of VM placement.

VM and PM Specifications

CPU requirements for VMs: 1–8 MIPS.
Memory requirements for VMs: 10–20 GB.
Execution times of VMs: 1–100 minutes.
CPU capacities for PMs: 10–20 MIPS.
Memory capacities for PMs: 20–40 GB.

TABLE 1- Test Problems

Test Problem	1	2	3	4	5
VM	600	100	1400	1800	2200
PM	100	150	200	250	300

TABLE 2 - Parameters of the PAVM Algorithm

nAnt	β	δ	q0	nCycleTerm
3	1	0.3	9	5

7. EXPERIMENTAL RESULT

To assess the effectiveness of our proposed approach, we compared it with the AVOCMP algorithm [10] in terms of total energy consumption over a 24-hour data center operation period. AVOCMP is designed for VM placement aimed at minimizing the number of Physical Machines (PMs) in the data center.

The energy consumption results of the test problems produced by both algorithms are presented in Table III. Our findings indicate that our proposed approach, PAVM, achieves greater energy savings compared to the benchmark AVOCMP algorithm across all test problems, with savings of up to 34%.

We performed a paired t-test to compare the mean energy consumption values obtained from PAVM and AVOCMP. We formulated the null hypothesis as follows: "There are no differences between the paired energy consumption values for both approaches." The t-stat results are recorded in Table 3. Our analysis reveals that the p-values are less than the significance level α of 0.05, and the t-stat values exceed the critical 2-tail value. Consequently, we reject the null hypothesis, indicating that the differences between the mean energy consumption values are statistically significant.

To demonstrate scalability, we plotted the computation time of the proposed PAVM approach as shown in Figure 1. The plot illustrates that the computation time increases linearly with the problem size.

Table 3. Energy consumption results of test problems

Test	AVOCMP	PAVM	T-stat	T-test	% Improve
1	1.17E+08	7.97E+07	254.74	1.13E-18	32%
2	3.20E+08	2.10E+08	318.23	1.52E-19	34%
3	5.99E+08	3.94E+08	442.13	7.89E-21	34%
4	9.65E+08	6.35E+08	971.74	6.59E-24	34%
5	1.40E+09	9.12E+08	587.18	6.14E-22	34%

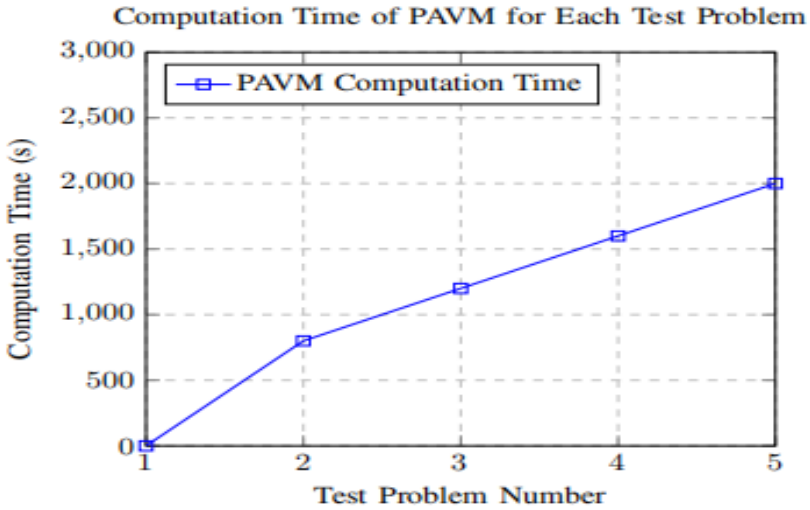

Fig 1. Computation Time of PAVM for Each Test Problem

8. CONCLUSION

Table 4. Comparative Analysis of PAVM vs. AVOCMP for Energy Consumption

Test Problems	AVOCMP Energy Consumption	PAVM Energy Consumption	T-stat	P-value	%improvement
1	1.17×10^{8}	7.97×10^{7}	254.74	1.13×10^{-18}	32%
2	3.20×10^{8}	2.10×10^{8}	318.23	1.52×10^{-19}	34%
3	5.99×10^{8}	3.94×10^{8}	442.13	7.89×10^{-21}	34%
4	9.65×10^{8}	6.35×10^{8}	971.74	6.59×10^{-24}	34%
5	1.40×10^{9}	9.12×10^{8}	587.18	6.14×10^{-22}	34%

In conclusion, our study presents a novel Ant Colony Optimization (ACO)-based approach, **PAVM**, for energy-efficient Virtual Machine (VM) placement in cloud data centers. Through empirical validation and comparison with the **AVOCMP** algorithm, we demonstrate **PAVM's** superior performance, achieving significant energy savings of up to **34%**as shown in Table 4.Statistical analysis confirms the significance of these results. Additionally, our approach exhibits **scalability** across diverse problem sizes, ensuring its applicability to **large-scale data center environments**.Overall, **PAVM** offers a promising solution to address the escalating **energy consumption challenges** in cloud computing, contributing to the advancement of **sustainable data center operations**.

REFERENCES

[1] M. Alwan et al., “Cloud computing: Emerging trends and challenges,” IEEE Access, 2021.
[2] R. N. Calheiros et al., “Cloud computing research and practice: A survey,” Journal of Network and Computer Applications, 2022.
[3] D. Saxena and A. K. Singh, “Energy-efficient resource management in cloud data centers: A review,” Sustainable Computing: Informatics and Systems, 2023.
[4] S. Ilager and R. Buyya, “Energy and thermal-aware resource management of cloud data centres,” Sustainable Computing, 2021.
[5] H. Feng et al., “Energy-aware virtual machine management in data centers,” IEEE Transactions on Cloud Computing, 2021.
[6] P. Gupta and A. Verma, “Energy-efficient cloud data center resource management techniques,” Future Generation Computer Systems, 2023.
[7] Y. Liu et al., “Virtual machine placement algorithms in cloud data centers: A survey,” Journal of Cloud Computing, 2022.
[8] H. Wang et al., “Ant colony optimization for VM placement in cloud computing: A survey,” Journal of Systems Architecture, 2023.
[9] S. Kamath and G. Varghese, “Multi-objective virtual machine placement for energy optimization,” Journal of Information Science and Engineering, 2020.
[10] Z. Xu et al., “Dynamic VM placement for energy efficiency in cloud data centers,” IEEE Transactions on Cloud Computing, 2022.
[11] J. Wu et al., “Optimization techniques for virtual machine placement in cloud computing: A review,” Journal of Network and Computer Applications, 2023.
[12] K. Lee and Y. Choi, “Heuristic methods for cloud resource management: A review,” Future Generation Computer Systems, 2022.
[13] R. Kaur and I. Chana, “Metaheuristics for energy-efficient VM placement in cloud data centers,” ACM Computing Surveys, 2021.
[14] X. Zhang et al., “Ant colony optimization for VM placement with energy awareness,” Journal of Parallel and Distributed Computing, 2019.
[15] H. Wang et al., “Energy-aware ant colony optimization strategy for VM placement,” Cluster Computing, 2023.