AIRO JOURNALS
Peer Reviewed Multidisciplinary International

ISSN:2320-3714
Volume:3 Issue: 2
August 2024
Impact Factor:10.2
Subject: Computer Science

OPTIMIZED CLOUD COMPUTING IOT RESOURCE ALLOCATION WITH AN ENHANCED ANT COLONY APPROACH

Jeelan Basha G
Associate professor
Harsha institute of management studies, Nelmagala, Bangalore

Abstract

The Internet of Things (IoT) is a disruptive force in the modern computing era because to its application of automation, superior ecosystems, innovative and effective services, and increased productivity. IoT has found uses in a number of domains, including industry, education, healthcare, agriculture, and the military, where a variety of resource requirements pose significant challenges. To address the difficulties of using genetic algorithms, expanding the full-scale task execution time, and low client fulfilment and resource use in view of current computations, a better ant colony calculation advancement method for cloud computing resource dissemination is planned considering the flexible Internet of things project. This will allow for easier completion of cloud computing asset allocation. Assignments in the portable Internet of things designing climate are ordered in light of cloud computing allocation features. Considering this, the genetic algorithm worldwide hunt capacity is added to the underlying allocation process to join it with the ant colony algorithm and use it during the time spent distributing cloud computing resources. The findings from the recreation demonstrate how the suggested method can shorten the time it takes to complete a task while also increasing customer satisfaction and efficiency.

Keywords: *Cloud Computing, IoT (Internet of Things), Resource Allocation, Optimized Resource Management, Ant Colony Optimization*

1. INTRODUCTION

The Internet of Things (IoT) facilitates the seamless integration of physical and digital environments by enabling various objects to connect with one another autonomously. The

ISSN:2320-3714
Volume:3 Issue:2
August 2024
Impact Factor:10.2
Subject:Computer Science

performance of IoT gadgets is impacted by a large number of factors and resources, for example, client needs, energy utilization, storage limit, correspondence needs, handling power, network bandwidth, and a range of uses. Because of their varied nature, these resources have a range of qualities and capacities. Resource allocation is an issue for Internet of Things networks, especially in networks with heterogeneous features. In order to accomplish optimal goals, resource allocation entails managing and allocating scarce resources in an efficient manner. Node resources and channel resources are the two types of resources found in Internet of Things networks. Storage, processing power, and energy resources are examples of node resources, sometimes referred to as physical resources. Conversely, channel resources are related to networks and communication channels and include things like load balancing, traffic analysis, and channel bandwidth.

Thinking about unambiguous factors and restrictions, resource allocation is the demonstration of proficiently dispersing accessible resources to finish a bunch of undertakings. The goal is to maximize the use of resources while improving the IoT platform's performance. IoT devices' resources are frequently constrained by things like processing speed, storage capacity, and energy limitations. IoT devices, however, have the ability to offer a variety of features and services. To make the best use of these scarce resources and guarantee that activities are executed successfully, efficient resource allocation is essential. Due to the dispersed and diverse nature of IoT devices and resources, allocating them becomes more complicated. There are many distinct kinds of IoT devices, each with unique capabilities and features. Their levels of energy, processing power, and storage capacity might differ. These variations must be taken into account and resources must be distributed appropriately using the resource allocation method.

Cloud computing is a relatively new technology that provides online storage and processing resources as well as the ability to structure applications with improved fault tolerance, availability, and scalability. The practice of storing data on distant servers rather than personal computers or other devices is known as cloud computing. Any device with cloud computing capabilities can access this information via the internet, no matter where in the globe it is located. The client side of the cloud computing system is called the front-end, and the data is stored on a collection of third-party servers and PCs that make up the back-end. A central server, a portion of the back end, facilitates communication between networked computers by using middleware and adhering to protocols. All of the computer resources are gathered and automatically managed by cloud computing. Its features—on-demand self-service, resource pooling, internet connectivity, elastic service availability, and tracking of services used by specific users—define a cloud computing system. Whether it's the replacement of Microsoft Office with Google Drive or traditional enterprise data storage with Amazon Web Services or online banking with branch offices, cloud computing is quickly becoming the norm. Services like Dropbox store all of our files and data. Even multiple deployment and service methods are offered by the cloud.

2. LITERATURE REVIEW

Wang, et al. [2023] fixated on tackling the appropriated asset portion issue in energy-effective information sending for Industrial Internet of Things (IIoT) frameworks with asset limitations. At this stage, they were approaching the test by representing the decentralized and somewhat obvious component of the framework as a Dec-POMDP. In order to address this issue and allocate assets in an energy-efficient manner, they promoted the innovative DADR approach of Dual Attention Assisted Deep Reinforcement Learning. The DADR computation makes use of a twofold consideration assisted profound support learning (DRL) model within the CTDE system, which stands for Convolutional Time, Attention, Dual-Attention, and Experience Reconstruction. As part of the DADR technique's entertainer organization, a multi-scale convolutional attention module (CAM) is used to extract highlight data from neighbourhood states in certain respects. We introduce an additional pundit network that delivers a thorough and accurate assessment of the framework state from a broad viewpoint by combining an encounter remaking module with a twofold consideration module. This pundit network jam adaptability in unique circumstances and proficiently handles non-fixed and to some extent recognizable issues in multi-specialist frameworks without requiring changes to the model design. The DADR calculation joins Multi-Head Self-Association (MHSA) with CAM to further develop the DRL model's portrayal learning execution.

Deng, et al. [2020] tackle the issue of overseeing dependability in edge computing (EC) frameworks, which are fundamental for dealing with the developing number of Internet of Things (IoT) gadgets that are associated with network edges. Their primary goal is to guarantee adherence to support level arrangements (SLAs), which are pivotal markers of the unwavering quality of Internet of Things (IoT) administrations. The creators propose utilizing a Markov

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume:3 Issue:2
August 2024
Impact Factor:10.2
Subject:Computer Science

choice cycle (MDP) to address the asset distribution plot and the condition of the help provisioning framework to resolve this issue. They utilize the level of SLA consistence as an intermediary for the dependability gain, which they encode and use as the objective for asset portion changes. To determine the best asset designation strategy, the creators use methods from Reinforcement learning (RL). They use support learning methods to prepare a strategy, permitting asset designation plans to be progressively produced in view of the present status of the framework. The preparation's strategy will likely distribute assets in a manner that boosts the administrations' development in believability. Tests are completed on the YouTube demand dataset to survey the proposed technique. The discoveries show that the edge administration provisioning framework utilizing the proposed strategy performs somewhere around 21.72% better than standard techniques.

Liu [2021] recommended a calculation for assigning assets to versatile edge computing. By considering various factors, including task appearance qualities, dormancy, energy utilization, link channel clog, and correspondence costs, the calculation tries to augment base station execution long term. We give a Lyapunov float punishment-based energy utilization setback line. To guarantee that energy utilization limits are fulfilled during improvement, this line relates the significant investment utilization of little base stations. The authors provide an offloading weight equation based on the Shapley value and use game theory to determine the offloading weight for the work phase. The offloading weight is decided in an unbiased manner, and it represents the arrival of different positions. The exhibition of the proposed calculation was evaluated utilizing MATLAB stage reproductions. The outcomes demonstrate the way that the technique can arrive at Nash harmony in a limited measure of cycles. Also, the calculation performs better compared to other examination calculations as far as energy utilization, time delay, and the quantity of errands that are effectively offloaded.

Nguyen, et al. [2020] give a summed up federated learning (FL) way to deal with tackle the issues with FL, like heterogeneous user Equipment (UE) and non-free and indistinguishably scattered information. Their procedure is to bring down the kind of worldwide correspondence while expanding FL's speed of combination. The recommended FL method adds a weight-based proximal term to the neighbourhood misfortune capability, developing the cutting-edge combined averaging (FedAvg) move toward currently being used. With this improvement, the calculation may proficiently bring down the correspondence above by running stochastic slope

AIRO JOURNALS
Peer Reviewed
Multidisciplinary
International
ISSN:2320-3714
Volume:3 Issue: 2
August 2024
Impact Factor:10.2
Subject: Computer Science

drop in lined up on an examined subset of UEs during each worldwide round. A combination upper bound that shows the compromise between the quantity of worldwide rounds and the intermingling rate is given by the specialists. The examination shows that even with a low number of dynamic UEs each round, union can in any case be guaranteed.

3. RESEARCH METHODOLOGY

3.1.Task Classification

Resource computing and resource storage are the two primary components of cloud computing, a novel kind of computer technology. The data center for cloud computing is mostly made up of servers and software resources, and it is responsible for supplying users with computing, storage, and other associated services.

A successful strategy for assigning resources presently needs to consider the virtual machine's heap and organization properties to actually take advantage of cloud computing resources, further develop resource use, and diminish task execution times.

Most frequently, cloud computing utilizes QoS allots to find how cheerful various clients are with the administrations they get. During the genuine allocation process, resources ought to be portrayed using QoS qualities so they can be disseminated all the more properly founded on various requirements. The Qos boundary can take on the accompanying shapes:

- **RAM:** If the user requires a large amount of memory on the computer for calculations, then the RAM parameters of Qos must be prioritized.
- **Network bandwidth:** Multimedia is needed for data transfer when network capacity has a bigger influence on users, and QoS network bandwidth criteria are given priority.
- **Cost:** Priority must be given to the cost aspects of QoS when the pricing issue affects users more.
- **Completion time:** To guarantee the task is completed quickly, users' requirements about the task completion time should be carefully taken into account while determining the QoS completion time criteria.
- **Network delay:** In the actual application process, network latency must be avoided if users have high real-time needs. To better serve the demands of various users, it is currently necessary to carefully analyze the network latency characteristics of Qos.

Based on the QoS metrics' previously mentioned classification, more satisfying services can be provided that cater to various client wants. In the interim, the job classification can be efficiently achieved by fusing QoS criteria with the features of cloud computing resources.

3.2.An Upgraded Ant Colony Algorithm for Allocating Resources in Cloud Computing

Rather than other transformative models, the ant colony calculation is a kind of irregular hunt that tracks down sensibly ideal responses by using a few haphazardly produced gatherings of initial solutions. The transformation stage and the coordination stage are the two particular stages that make up the whole interaction. During the transformation stage, each arrangement is adjusted fundamentally by modifying the data associated with the advancement interaction; during the coordination stage, each arrangement utilizes a data trade system to get the arrangement that is similarly ideal.

The quantity of urban areas is set to n, and the quantity of ants to m. Everything begins with the accompanying arrangements, according to the unique habits of each ant:

- When ants are deciding which city to visit next, they use the correlation probability function, distance to known city i, and path pheromone ηβij(t) as their basis: ταij(t).
- Ants choose their paths based on pre-established guidelines, most of which are governed by the taboo table. Ants that have not gone through a cycle are unable to go to other cities.
- Since ants leave behind a specific quantity of pheromone along their path during migration, the probability function is given as

$$p_k^{ij} = \frac{\tau_{ij}^{\alpha}(t) \cdot \eta_{ij}^{\beta}(t)}{d_k}.$$

Following the completion of the cycle by every ant, the following formula updates every pheromone along the path:

$$\tau_{ij}(t+n) = (1-p) \times \tau_{ij}(t) \times p_k^{ij},$$

where t is the movement time of the ants. However, the following issues will arise when ant colonies are really used:

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume:3 Issue: 2
August 2024
Impact Factor:10.2
Subject: Computer Science

- Low initial solution speed
- Calculation process prone to stalling
- Reduced efficiency at large system scales

Here, the genetic algorithm is integrated into the ant colony method due to its fast and arbitrary global search capabilities. The correctness of the original algorithm is preserved while the convergence speed is substantially enhanced by integrating these two methods. In order to prevent the algorithm from becoming stuck in a local optimal solution and to maintain or even improve species diversity, the technique of reverse transformation is also introduced.

There is some blindness in the early search stage of the ant colony algorithm since there is less pheromone accumulation on each path and there is a tiny change during the first iteration. On the other hand, the ant colony method's initial information distribution is based on the genetic algorithm's better solutions since it can rapidly explore the environment and is used for both allocation and search. This makes it possible to swiftly acquire the ant colony algorithm's initial information distribution, which effectively makes up for the algorithm's early deficiencies in search capability and accelerates the algorithm's convergence. After the initial pheromone conveyance is created using the genetic algorithm, the ant colony algorithm is utilized for looking.

The fitness function, a particular evaluation criterion, is used in the genetic algorithm to assess each member of the population's quality. The precise formula for calculation is

$$f(i) = \left[\frac{\sum_{k=1}^{M} (t\,\mathrm{time}\,(k) - \min + 1)}{M \cdot (t\,\mathrm{time}\,(k) - \min + 1)}\right]^2,$$

where M is the populace size and time (k) is the greatest fulfilment season of the ith task allocation system. "min" signifies the absolute minimum of time expected to follow through with each job in the work allocation plan.

Genetic algorithms can be coded using a variety of techniques, and the following are some uses of real number coding technology. A set of guidelines is used to pick the chromosomes with superior performance when the initial population is formed. The process of creating new individuals in genetic algorithms relies heavily on mutation and crossover processes.

Following this, the pheromone transformation factor is used to turn the ant population at each resource point into the underlying pheromone on the asset point's pheromone pathway. In the real world, the ant colony algorithm is built around the arrangement obtained from the genetic algorithm.

Below you can see the detailed steps of how this algorithm works:

a) Figure out the initial values of the ant colony's and genetic algorithms' parameters, which are mostly the population size and evolutionary algebra.
b) Form the initial population and set the initial evolution times to 0.
c) Increase every evolutionary algebraic term by one and carry out crossover, mutation, and copy operations.
d) Determine whether step
e) may complete the maximum number of iterations; proceed to step (e) if it satisfies the conditions, otherwise go to step (c).
f) Transform the genetic algorithm's result into the original pheromone.

Table 1: Adjusting the parameters of the experiment

Name of Parameters	Specific value
Ant Colony Iterations	900 Times
Maximum Iterations	40 Times
Minimum Iterations	30 Times
Crossover Probability	0.9%
Population size	50
Mutation probability	0.06
Minimum evolution rate	6%
Hardware configuration	400G hard disk, 8G RAM
Operating system	Windows 10
Additional RAM	512G

In Table 1 you can see the experimental parameters for an optimization technique for ant colonies. With a minimum of 30 and a maximum of 40 iterations, the algorithm is configured to ensure convergence. The ant colony will undergo nine hundred iterations in order to explore possible solutions. A preset mutation probability of 0.07 introduces random alterations, while a fixed crossover probability of 0.8% determines the possibility of merging solutions. The method has 50 populations, and the minimal evolution rate needed to advance is set at 6%.

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume:3 Issue: 2
August 2024
Impact Factor:10.2
Subject: Computer Science

Operating on Windows 10, the physical environment consists of a 400 GB hard drive and 8 GB of RAM. An extra 512 GB of RAM is listed, which probably indicates a redundant entry or typo.

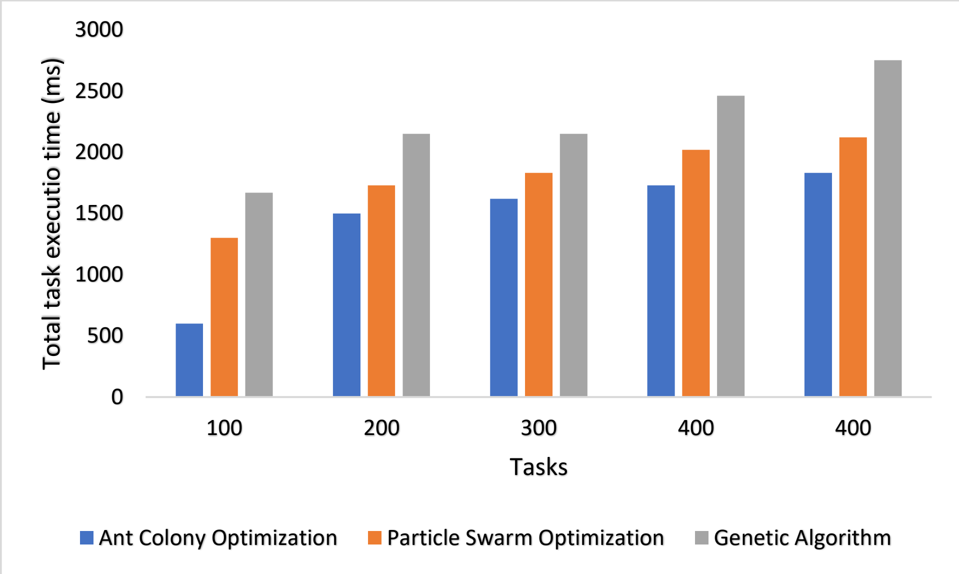

Figure 1: Comparison of the overall task execution time for each approach

g) Assign n nodes to ants or resources, or m resources to nodes.
h) Create a relationship between tasks and available resources; begin at the top of the taboo table; refine each ant; and proceed to the next vertex, j, to update the taboo table.
i) Modify all pheromone foci to check if the ant has finished a task. Once the errand is completed, proceed with the next step.
j) in that case, go on to step (g). I will revise the total pheromone strength.
k) Increase the iteration count by one and apply the mutation method to all pheromone concentrations.
l) Find out if one has reached the maximum number of iterations. If that's the case, go to step (g). Otherwise, go to step (l).
m) Determine the best possible result by applying the objective function to the current one.

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume:3 Issue: 2
August 2024
Impact Factor:10.2
Subject: Computer Science

4. DATA ANALYSIS

An experiment is conducted using MATLAB to verify that the modified ant colony approach for allocating resources in the cloud, which is based on the mobile Internet of things project, is generally effective.

The techniques in and are chosen for comparison, respectively. The primary metrics used to verify the methods' performance are task completion time, resource usage, and user satisfaction. Table 1 displays the experimental parameter settings.

4.1.Total Time of Task Execution.

Three distinct approaches were compared in terms of the total time required to complete the task (Figure 1).

Analyses of experimental data in Figure 1 reveal that the proposed method employs a total task execution time that is significantly lower than the other two approaches. Nevertheless, the disparity in the total execution time of each method becomes more pronounced as the quantity of tasks grows. The results of the simulations for different job counts make this quite clear; they reveal that the proposed strategy is not particularly beneficial for low work counts. This is mainly because the proposed method employs a genetic algorithm to improve upon the ant colony algorithm, rendering it more effective in handling intricate problems involving the distribution of resources.

4.2.User Satisfaction.

Table 2 compares the three approaches in terms of user satisfaction.

Table 2: Comparison of user satisfaction with different algorithms

No. of Experiments (times)	Ant Colony Optimization	Particle Swarm Optimization	Simulated Annealing (SA)
10	98.28	97.91	94.78
15	97.07	95.24	92.61
20	96.93	94.21	91.33
25	96.51	93.56	90.36
30	95.18	92.30	89.05
35	95.11	90.40	87.07
40	94.98	89.73	85.21
45	94.66	87.31	84.63
50	94.17	86.40	83.15

ISSN:2320-3714
Volume:3 Issue: 2
August 2024
Impact Factor:10.2
Subject: Computer Science

Table 2 compares user satisfaction for three optimization techniques that were tested across varying numbers of experiments: Particle Swarm Optimization (PSO), Simulated Annealing (SA), and Ant Colony Optimization (ACO). ACO satisfaction levels typically drop to 94.17% after 50 studies, having peaked at 98.28% after 10 experiments. PSO starts off marginally lower at 97.91% but likewise gradually declines, reaching 86.40% after 50 trials. SA exhibits the lowest initial satisfaction (94.78%), which declines more sharply over time to 83.15%. This pattern suggests that although ACO continuously maintains higher user satisfaction than PSO and SA, as the number of experiments increases, the contentment of users decreases for all three algorithms.

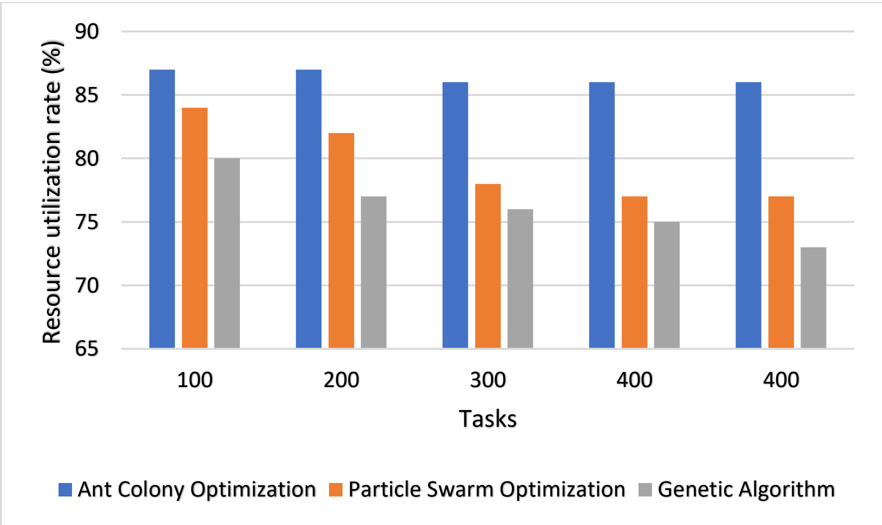

Figure 2: Comparing how different strategies use resources

4.3.Resource Utilization.

Three distinct algorithms' resource consumption is compared in Figure 2.

Experiment results shown in Figure 2 show that when the number of jobs increases, the Ant colony optimization method's resource usage gets better. This further supports the validity of the ant colony algorithm's incorporation of a genetic algorithm. Following enhancement, the entire method's resource usage rate is noticeably higher than that of the other two approaches.

5. CONCLUSION

The results of the experiments prove that the proposed method may maximize resource use, boost user enjoyment, and reduce task completion times. Due to the restricted amount of productive research time and energy, the suggested strategy still has certain shortcomings. Given the differences between experimental and real data, testing and improving the suggested method requires applying it to real-world contexts. More parameter optimization is required to reduce the effect on the entire procedure. To keep up with the ever-evolving demands of cloud computing, future studies should employ more complex algorithms. In addition, maintaining efficient use of resources depends on generic task model optimization.

REFERENCES

1. *Chen, L. Wang, and J. Ren, "Multi-constraint cloud workflow scheduling algorithm based on resource grouping," Journal of University of Electronic Science and Technology of China, vol. 46, no. 3, pp. 562–568, 2017.*
2. *D. Jiang and H. Lin, "A review of key technologies for resource allocation in cloud computing environments," Journal of China Academy of Electronics, vol. 13, no. 3, pp. 308–314, 2018.*
3. *G. Sun, "Virtual machine dynamic supply and allocation algorithm based on auction in cloud computing," Computer Engineering and Design, vol. 38, no. 6, pp. 1547–1550+1555, 2017.*
4. *H. Wang, Z. Li, X. Zhang, H. Bu, L. Kong, and X. Zhao, "Virtual machine resource allocation method based on historical data," Journal of Computer Research and Development, vol. 56, no. 4, pp. 105–115, 2019.*
5. *J. Cui, J. Liu, and C. Yang, "Network delay cloud computing resource scheduling algorithm based on priority queuing theory," Journal of Guilin University of Technology, vol. 37, no. 2, pp. 360–365, 2017.*
6. *J. Liu, "Task offloading and resource allocation algorithm based on mobile edge computing in Internet of Things environment," The Journal of Engineering, vol. 2021, no. 9, pp. 500-509, 2021.*
7. *L. Cai, "Research on cloud computing resource allocation based on particle swarm optimization," Journal of South China Normal University Natural Science Edition, vol. 42, no. 9, pp. 128–132, 2017.*

8. *S. Deng et al., "Dynamical resource allocation in edge for trustable Internet-of-Things systems: A reinforcement learning method," IEEE Transactions on Industrial Informatics, vol. 16, no. 9, pp. 6103-6113, 2020.*
9. *V.-D. Nguyen, S. K. Sharma, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Efficient federated learning algorithm for resource allocation in wireless IoT networks," IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3394-3409, 2020.*
10. *X. Li, C. Qian, and X. Liu, "Resource allocation mechanism of cloud workflow system based on dynamic pricing combination reverse auction," Computer Integrated Manufacturing Systems, vol. 23, no. 05, pp. 941–946, 2017.*
11. *X. Luo, L. Wang, and G. Xin, "Analysis of influencing factors of virtual machine allocation performance in cloud computing environment," Journal of Computer Applications, vol. 37, no. S2, pp. 29–31+48, 2017.*
12. *Y. Wang, F. Shang, J. Lei, X. Zhu, H. Qin, and J. Wen, "Dual-attention assisted deep reinforcement learning algorithm for energy-efficient resource allocation in Industrial Internet of Things," Future Generation Computer Systems, vol. 142, pp. 150-164, 2023.*
13. *Y. Wang, J. Wang, and X. Song, "Improvement of cloud computing resource Nash equilibrium optimization allocation method," Computer Engineering, vol. 43, no. 12, pp. 17–24, 2017.*
14. *Y. Zheng, C. Song, and W. Zhao, "Multi-level fairness QoS constraint task scheduling algorithm in cloud computing," Journal of Shenyang University of Technology, vol. 41, no. 3, pp. 309–314, 2019.*
15. *Z. Liu and J. Zhou, "Load balancing dominant resource fair allocation algorithm," Computer Engineering and Science, vol. 41, no. 9, pp. 1574–1580, 2019.*

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume:3 Issue: 2
August 2024
Impact Factor:10.2
Subject: Computer Science

Author's Declaration

I as an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my whole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation/Address of my university/college/institution/ Structure or Formatting/ Resubmission/Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible or shuffled or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who finds trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide ordo not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher then my paper maybe rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds Any complication or error or anything hidden or implemented otherwise, my paper may be removed from the website or the watermark of remark/actuality maybe mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Jeelan Basha G

******