STM JOURNALS
Journal of
Computer Technology & Applications
https://journals.stmjournals.com/jocta

ISSN: 2229-6964 (Online)
ISSN: 2347-7229 (Print)
Volume 17, Issue 1, 2026
January–April
DOI (Journal): 10.37591/JoCTA

Green and Edge-Aware Computing: Rethinking Cloud Infrastructure for Sustainability

Dipankar Barui[1,]*, Raghunath Maji[2], Subhankar Roy[3], Satyajit Maiti[4], Subhadip Goswami[5]

Abstract

Cloud computing has transformed the way organizations access and manage information technology resources, providing flexible, scalable, and cost-efficient services that support today's data-driven world. Despite these advantages, the rapid expansion of large-scale cloud infrastructures has resulted in rising energy consumption, significant heat generation, and a growing environmental footprint. This research focuses on advancing green cloud computing by examining methods that reduce power usage while maintaining high performance. Key strategies include improved virtualization techniques, smarter resource scheduling, and the use of thermal energy recovery systems to repurpose excess heat. In addition, the study highlights the increasing relevance of edge computing as a complementary approach to achieving sustainability. By processing data closer to end devices, edge systems help decrease latency, lower network congestion, and reduce reliance on energy-intensive centralized data centers. To further enhance efficiency, the paper explores various load balancing methods, with particular attention to greedy-based algorithms designed to distribute workload intelligently across edge servers. These techniques support optimal resource allocation, minimize processing delays, and ensure smoother network operations, contributing to a more sustainable and responsive computing environment.

Keywords: Cloud-Fog computing, green computing, Internet of Things (IoT), sustainability, virtualization

INTRODUCTION

In today's fast-paced digital world, the demand for smart and sustainable technologies is critical. Cloud computing, a key enabler of modern digital infrastructure, offers numerous advantages, including scalability, high availability, and global accessibility. However, its rapid expansion has led to significant environmental concerns owing to the high-power consumption and CO_2 emissions from data centers [1]. These centers require continuous operation and cooling, which results in energy loss and environmental degradation. Consequently, there is a growing focus on adopting green cloud computing, which emphasizes energy-efficient resource utilization, recycling, and reducing the ecological footprint of information technology services. Green computing promotes eco-friendly practices such as offloading processing loads and implementing heat recovery methods. To address the challenges posed by the cloud infrastructure, the shift toward edge computing is gaining momentum. Edge computing processes data closer to its source, offering benefits such as reduced latency, lower energy use, and a faster real-time response. This

***Author for Correspondence**
Dipankar Barui
E-mail: dipankarbarui123@gmail.com

[1]Assistant Professor, Department of Artificial Intelligence and Machine Learning, St. Thomas' College of Engineering and Technology, Kolkata, West Bengal, India
[2]Assistant Professor, Department of Computer Science and Engineering, Greater Kolkata College of Engineering and Management, Kolkata, West Bengal, India
[3]Assistant Professor, Department of Computer Science and Engineering, Seacom Engineering College, Kolkata, West Bengal, India
[4]Assistant Professor, Department of Computer Science and Engineering, Swami Vivekananda University, Kolkata, West Bengal, India
[5]Assistant Professor, Department of Computer Science and Technology, JIS College of Engineering, Kolkata, West Bengal, India

Received Date: October 14, 2025
Accepted Date: December 12, 2025
Published Date: February 20, 2026

Citation: Dipankar Barui, Raghunath Maji, Subhankar Roy, Satyajit Maiti, Subhadip Goswami. Green and Edge-Aware Computing: Rethinking Cloud Infrastructure for Sustainability. Journal of Computer Technology & Applications. 2026; 17(1): 17–24p.

model is well-suited for the rise in IoT devices and 5G networks, which demand rapid data processing. By distributing computational loads more efficiently, edge computing supports both the performance and sustainability goals. Gartner predicted that by 2025, 75% of enterprise data processing will take place at the edge, emphasizing its critical role in advancing sustainable and smart computing environments [2]. As reported by DataReportal [3], 63.5% of the world's population is now online, and the rapid rise in internet usage has accelerated the emergence of smart technologies designed for real-time interaction. Figure 1 illustrates data processing across different layers within the edge computing architecture.

Towards Green Computing: Literature Survey

Tyagi et al. [4] focused on green cloud computing, emphasizing energy-efficient practices and renewable energy integration. Although it does not specifically address edge computing, the principles of sustainability and resource optimization can be applied to both paradigms for enhanced environmental benefits.

Dash et al. [5] emphasized the importance of green computing in minimizing the environmental harm caused by waste, energy use, and hazardous materials. It highlights how mobile technologies such as IoT, fog, mist, edge, and cloud can be made energy-efficient while also addressing related research challenges in achieving sustainable computing.

Heidari et al. [6] proposed an energy-efficient IoT-edge offloading system using deep learning, Markov decision process, and blockchain for secure, low-latency processing. It introduces a deep Post-Decision State-learning algorithm to improve the learning speed and performance, achieving notable reductions in delay, energy use, cost, and computational overhead compared to existing methods.

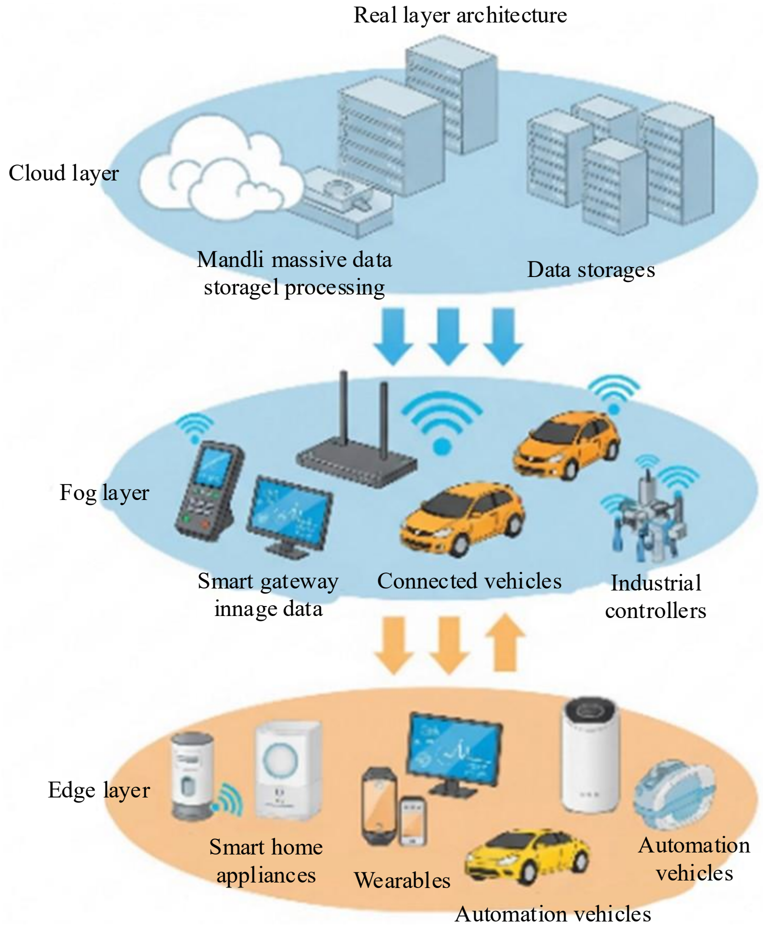


Figure 1. Three-layer architecture of cloud, fog, and edge computing.

Haseeb et al. [7] introduce an "intelligent and secure edge-enabled computing (ISEC) model" designed to enhance the monitoring and operations of sustainable cities, leveraging green Internet of Things (IoT) principles. The core motivation for this model stems from the inherent limitations of current IoT systems, particularly concerning energy management, data security, and communication reliability for low-power sensors in large-scale deployments.

Cao et al. [8] presented an in-depth survey on the integration of edge and edge-cloud computing within cyber-physical systems (CPS), with a strong emphasis on improving the quality of service (QoS). It explores the complexities of handling massive data volumes in CPS and demonstrates how edge-based approaches can significantly enhance the system's efficiency and responsiveness.

Ma et al. [9] investigated energy efficiency and task offloading in edge computing; however, *Green Edge* offered a distinct, integrated solution by coordinating green energy scheduling with dynamic task offloading in multi-tier edge systems, focusing on Multi-access Edge Computing (MEC) server optimization and effectively handling the variability of green energy across hierarchical infrastructures.

Khanh et al. [10] proposed an advanced edge computing management mechanism tailored for IoT-enabled smart cities, leveraging a lightweight "information map" to facilitate the seamless exchange of service-related data between edge servers as users transition across coverage areas. Their approach significantly enhanced response time and reduced energy consumption, promoting efficient and sustainable urban computing environments.

Mohammadzadeh et al. [11] highlighted the challenges of workflow scheduling in cloud environments, emphasizing the need for advanced heuristic and metaheuristic algorithms. This study stresses the importance of balancing exploration and exploitation, avoiding local optima, and addressing multi-objective goals, particularly energy efficiency, to advance green cloud computing.

Rahimikhanghah et al. [12] examined the limitations of cloud computing, such as latency and suboptimal resource utilization, and highlighted fog computing as a solution that extends services closer to end users.

CONCEPT

Green Cloud Computing is Evolving as a Cutting-Edge Innovation

Green cloud computing is an environmentally conscious approach to the design, development, and management of cloud infrastructure, with the goal of reducing its ecological footprint. It emphasizes energy-efficient operations by maximizing computational output while minimizing resource consumption. This method not only reduces the use of hazardous materials but also improves energy utilization throughout the lifecycle of cloud infrastructure. By lowering the power demands, cooling requirements, and operational costs, green cloud computing provides a more sustainable and cost-effective alternative to traditional cloud models. It empowers businesses to adopt greener practices while continuing to benefit from the scalability and flexibility of cloud services. The ultimate aim is to reduce the substantial energy consumption of data centers operated by major providers, such as Amazon Web Services (AWS), Azure, and Google Cloud Platform (GCP), thereby supporting global efforts toward a sustainable digital future.

Various approaches to implementing green computing methods.

1. *Virtualization and resource consolidation*: This enables multiple virtual machines (VMs) to run on a single physical server, maximizes resource utilization, and minimizes idle energy consumption.
2. *Dynamic resource allocation and load balancing*: Allocates computing resources based on real-time demand, avoiding over-provisioning and reducing power waste. Intelligent load balancing ensures optimal server usage.

3. *Energy-efficient data center design*: This includes energy-saving hardware, efficient cooling systems (e.g., liquid cooling and free cooling), and architectural designs that reduce heat and electricity consumption.
4. *Task scheduling based on energy awareness*: scheduling tasks at off-peak hours or directing them to low-energy-consuming servers, thereby optimizing energy distribution across the cloud infrastructure.

Carbon-Aware Computing: Cloud providers implement carbon-tracking tools and migrate workloads to locations or times when energy sources are greener (e.g., using carbon-intensity APIs).

Applications of Green Cloud Computing

According to Zhang et al. (2021) [13], the current research landscape primarily focuses on the following key areas:

- Energy-efficient data center management
- Sustainable big data analytics
- Eco-friendly cloud-based IoT systems
- Green artificial intelligence (AI) and machine learning model training
- Cloud-based smart city solutions
- Energy-aware virtualization techniques
- Green Service Level Agreement (SLA) and QoS research
- Carbon-aware cloud workload scheduling
- Hybrid edge-cloud models for green computing
- Blockchain for green cloud governance

Advantages of Green Cloud Computing

- Energy efficiency
- Reduced carbon emissions
- Cost savings
- Improved resource utilization
- Extended equipment lifespan
- Support for sustainable information technology practices
- Regulatory compliance
- Positive brand image
- Scalability with sustainability
- Promotion of renewable energy use
- Enhance server utility rate

Eco-Efficient Edge Computing via Green Information Technology Strategies

The rise in digital service demand has significantly impacted global energy consumption, with data centers and transmission networks contributing approximately 1% of energy-related greenhouse gas (GHG) emissions. A 2019 survey revealed that just 30 minutes of Netflix streaming could emit approximately 18 g of CO_2. As internet usage has surged—doubling since 2010 and increasing web traffic twentyfold—concerns about the environmental footprint of services, such as video streaming, have intensified. Reports such as those from the Shift Project estimate that Netflix's streaming may consume approximately 94 TWh annually, far exceeding Netflix's 2019 claim of 0.45 TWh. To address these concerns, enhancing the sustainability of data centers has become a priority. Strategies include reducing emissions, maximizing resource efficiency, and minimizing waste. Green cloud computing is gaining prominence, with over 77% of users implementing or planning to use energy-efficient UPS systems in edge networks. These modern UPS systems achieve up to 98% operational efficiency through dynamic online optimization, reducing power losses and ensuring reliable switching during power fluctuations. Additionally, there is a growing adoption of sustainable technologies, such as renewable energy sources, dynamic grid support, and low-GWP solutions, particularly in edge computing environments, where managing heat and power efficiency is becoming increasingly important.

PROPOSED METHODOLOGY

Basic Concept of Cloud Load Balancing

According to Rahman et al. (2014) [14], load balancing is a crucial technique in cloud computing that is used to efficiently distribute incoming client requests among multiple servers to enhance performance and response time (Figure 2). A typical architecture involves a load balancer that receives requests, estimates their load, monitors the server status, and applies appropriate algorithms to assign tasks. This ensures high resource utilization, traffic control, and system reliability without overloading a single server. Effective load balance supports scalability, improves availability, and reduces infrastructure cost. The key features include minimal overhead, current system load awareness, periodic balancing, efficient resource migration, and reliable communication. This can be implemented in a centralized or decentralized architecture.

Proposed Greedy Approach for Green Cloud Computing

The weighted least connection (WLC) algorithm is an intelligent load balancing strategy that enhances traditional least connection methods by assigning weights to servers based on their processing capabilities (Table 1). The WLC greedy-based approach greedily selects the server with the lowest ratio of active connections to its weight, ensuring an optimized and balanced allocation of requests in heterogeneous cloud environments.

Table 1. A Comparative study of cloud computing and edge computing based on green cloud.

Aspect	Cloud computing	Edge computing
Data processing location	Centralized (data centers)	Decentralized (near the data source)
Energy consumption	Higher, due to large-scale data centers	Lower, due to localized and distributed processing
Carbon emissions	High, due to heavy reliance on power-intensive infrastructure	Reduced by minimizing data travel and using lighter devices
Latency	Higher latency due to data traveling longer distances	Lower latency as data is processed near the source
Network bandwidth usage	Consumes more bandwidth	Optimizes bandwidth by reducing data transmission
Cooling requirements	Significant cooling is needed for large data centers	Minimal cooling required for smaller, distributed nodes
Power losses	More power is lost in transmission and conversion	Reduced losses due to proximity to end devices
Renewable energy integration	Gradual adoption in large facilities	Easier integration at the micro-scale with localized setups
Scalability in the green context	Limited by the centralized infrastructure's environmental cost	Highly scalable with energy-efficient local nodes
Sustainability focus	Emerging requires a major infrastructure overhaul	Naturally aligns with sustainable computing goals
Cost efficiency (green focus)	High operational cost due to cooling and energy	More cost-effective with optimized energy use
Environmental impact	Larger ecological footprint	Smaller footprint due to decentralized, efficient setup

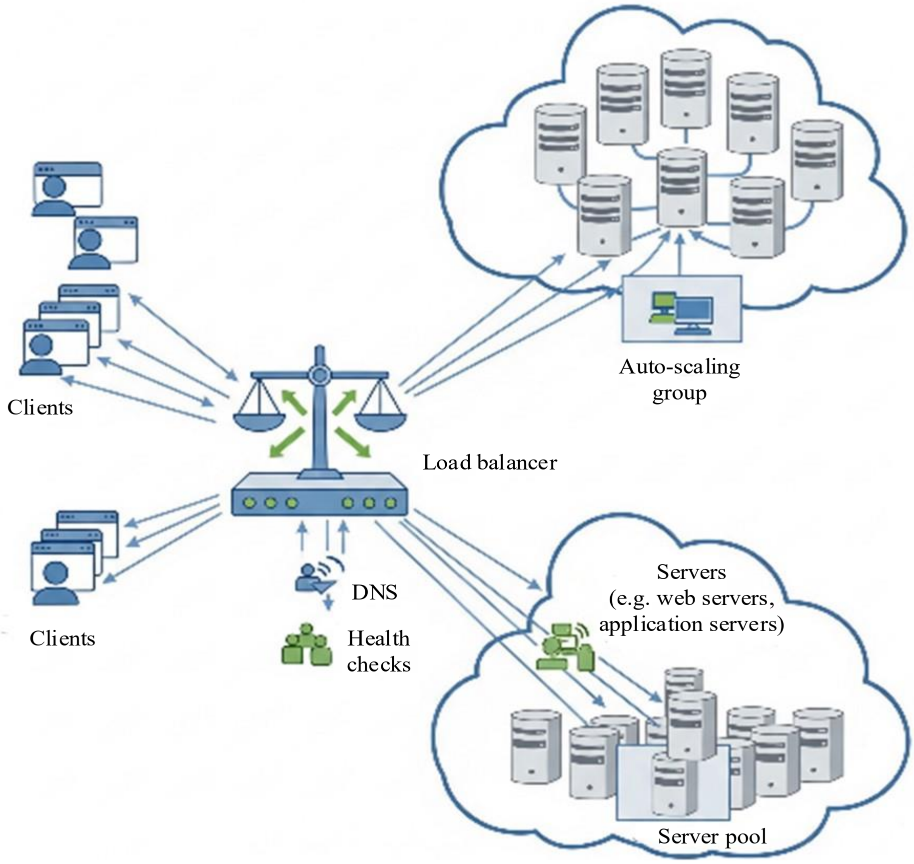


Figure 2. Load balancing in green cloud computing.

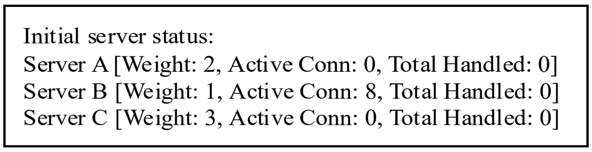
Algorithm: WLC Greedy Approach

1. Each server is assigned a weight according to its capability Central Processing Unit (CPU) power, memory, etc.
2. The algorithm calculates the load ratio for each server as:
   load ratio = current connections/weight.
3. The request is greedily assigned to the server with the minimum load ratio
4. This ensures that powerful servers handle more connections than weaker servers, thereby improving the overall throughput and resource utilization.

RESULT AND DISCUSSION

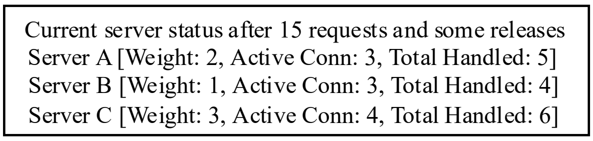
In Figures 3 and 4, the WLC greedy-based load balancer demonstrates dynamic request distribution in a cloud computing environment, ensuring optimal server utilization based on real-time load ratios and server capacities. Initially, a set of servers is created, and each is assigned a unique identifier and a specific weight reflecting its computational capability (such as CPU power or memory capacity). The algorithm calculates the load ratio for each server by dividing the number of active connections by the server's weight (active connections/weight). Incoming requests are greedily assigned to the server with the lowest load ratio, ensuring that higher-capacity servers handle more traffic than less-capable servers. Figure 3 shows the initial status of the simulator.

During execution, the algorithm processes an initial batch of 15 simulated requests periodically, releasing some to emulate real-world traffic behavior. The output includes detailed logs of request assignments, releases, and the evolving status of each server. To demonstrate scalability, a new server (*Server D*) is added, and five additional requests are handled, showing how the load balancer immediately integrates new resources. Subsequently, *Server B* was removed to simulate downtime or maintenance, and five additional requests were processed, confirming the adaptability of the algorithm. Figure 4 shows the final status of the simulator.



```
Initial server status:
Server A [Weight: 2, Active Conn: 0, Total Handled: 0]
Server B [Weight: 1, Active Conn: 8, Total Handled: 0]
Server C [Weight: 3, Active Conn: 0, Total Handled: 0]
```

Figure 3. Initial status of the simulator.



```
Current server status after 15 requests and some releases
 Server A [Weight: 2, Active Conn: 3, Total Handled: 5]
 Server B [Weight: 1, Active Conn: 3, Total Handled: 4]
 Server C [Weight: 3, Active Conn: 4, Total Handled: 6]
```

Figure 4. Final status of the simulator.

The final output summarizes the total requests handled by each server, illustrating how the WLC algorithm efficiently balances the load based on the server capacity while maintaining responsiveness, flexibility, and high availability.

CONCLUSION

In conclusion, the implemented WLC greedy-based load balancer demonstrates an efficient approach to distributing workloads across multiple servers based on their capabilities. By prioritizing servers with fewer weighted active connections, the algorithm ensures the optimal use of available resources, reduces the chances of server overload, and improves the overall response time. The dynamic handling of server additions and removals also demonstrates the flexibility of the system in adapting to real-time cloud infrastructure changes. This simulation highlights the practical application of WLC in achieving scalable and energy-efficient resource management in cloud computing environments.

Future Scope

The WLC greedy-based load balancing method supports green cloud computing by enhancing energy efficiency in expanding data centers. Future improvements may include energy-aware scheduling, AI-driven traffic prediction, VM migration to reduce power use, and carbon footprint tracking. Extending WLC to edge-fog cloud systems can further reduce energy consumption and promote an intelligent, eco-friendly cloud infrastructure.

REFERENCES

1. Garg SK, Yeo CS, Buyya R. Green cloud framework for improving carbon efficiency of clouds. In: Jeannot E, Namyst R, Roman J, editors. Euro-Par 2011 Parallel Processing. Lecture Notes in Computer Science. Vol. 6852. Berlin (DE): Springer; 2011. p. 491–502. doi:10.1007/978-3-642-23400-2_45.
2. Raval A. (2018). What edge computing means for Infrastructure and Operations leaders. [online] Express Computer. Available from: https://www.expresscomputer.in/news/what-edge-computing-means-for-infrastructure-and-operations-leaders/24531/
3. DataReportal. (2025). Digital 2026: Global Overview Report. [online] DataReportal – Global Digital Insights. Available from: https://datareportal.com/reports/digital-2026-global-overview-report
4. Tyagi AK, Cherian AK, Tiwari S. Green cloud computing: Opportunities and challenges. In: D L, Nagpal N, Kassarwani N, Varthanan GV, Siano P, editors. A sustainable future with E-Mobility: Concepts, Challenges, and Implementations. Hershey (PA): IGI Global; 2024. p. 226–252. doi:10.4018/979-8-3693-5247-2.ch012.
5. Atta-ur-Rahman, Dash S, Ahmad M, Iqbal T. Mobile cloud computing: A green perspective. In: Udgata SK, Sethi S, Srirama SN, editors. Intelligent Systems. Lecture Notes in Networks and Systems. Vol. 185. Singapore: Springer; 2021. p. 523–533. doi:10.1007/978-981-33-6081-5_46.
6. Heidari A, Navimipour NJ, Jamali MAJ, Akbarpour S. A green, secure, and deep intelligent method for dynamic IoT-edge-cloud offloading scenarios. Sustain Comput Inform Syst. 2023;38:100859. doi:10.1016/j.suscom.2023.100859.

7. Haseeb K, Ud Din IU, Almogren A, Ahmed I, Guizani M. Intelligent and secure edge-enabled computing model for sustainable cities using green Internet of Things. Sustain Cities Soc. 2021;68:102779. doi:10.1016/j.scs.2021.102779.
8. Cao K, Hu S, Shi Y, Colombo AW, Karnouskos S, Li X. A survey on edge and edge-cloud computing assisted cyber-physical systems. IEEE Trans Ind Inform. 2021;17(11):7806–7819. doi:10.1109/TII.2021.3073066.
9. Ma H, Huang P, Zhou Z, Zhang X, Chen X. GreenEdge: Joint green energy scheduling and dynamic task offloading in multi-tier edge computing systems. IEEE Trans Veh Technol. 2022;71(4):4322–4335. doi:10.1109/TVT.2022.3147027.
10. Vu Khanh Q, Nguyen VH, Minh QN, Dang Van A, Le Anh N, Chehri A. An efficient edge computing management mechanism for sustainable smart cities. Sustain Comput Inform Syst. 2023;38:100867. doi:10.1016/j.suscom.2023.100867.
11. Mohammadzadeh A, Masdari M, Gharehchopogh FS, Jafarian A. Improved chaotic binary grey wolf optimization algorithm for workflow scheduling in green cloud computing. Evol Intell. 2021;14(4):1997–2025. doi:10.1007/s12065-020-00479-5.
12. Rahimikhanghah A, Tajkey M, Rezazadeh B, Rahmani AM. Resource scheduling methods in cloud and fog computing environments: A systematic literature review. Cluster Comput. 2022;25(2):911–945. doi:10.1007/s10586-021-03467-1.
13. Zhang Y, Zhang CH, Shao X. User preference-aware navigation for mobile robot in domestic via defined virtual area. J Netw Comput Appl. 2021;173:102885. doi:10.1016/j.jnca.2020.102885.
14. Rahman M, Iqbal S, Gao J. Load balancer as a service in cloud computing. 2014 IEEE 8th International Symposium on Service Oriented System Engineering, Oxford, UK. 2014. p. 204–211. doi:10.1109/SOSE.2014.31.