Power System Technology
ISSN:1000-3673

Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

Resource Allocation and Scheduling in Fog–Cloud Computing: A Comprehensive Survey of Metaheuristic, Learning-Based, and QoS-Aware Approaches

[1]Ravi Kumar S G,
Assistant Professor, Dept. of CSE, R N S Institute of Technology, Bangalore-India
sgravikumar@gmail.com

[2]Dr. Anjan Kumar K N,
Associate Professor, Dept. of CSE, R N S Institute of Technology, Bengaluru- India
anjankn05@gmail.com

[3]Dr.Prasanna Kumar M,
Associate Professor, Dept. of CSE, R N S Institute of Technology, Bengaluru- India.
prasannamysoru@gmail.com

[4]Dr. Hemanth S,
Professor Dept. of CSE, R N S Institute of Technology, Bengaluru- India.
hemanth.shantha@gmail.com
Corresponding author: Dr. Anjan Kumar K N: anjankn05@gmail.com

Abstract:- Fog cloud computing has become a vital paradigm for next-generation networks and the Internet of Things (IoT), bringing computation and storage closer to the edge. This integration reduces latency, improves energy efficiency, and enhances quality of service (QoS). However, effective resource allocation and task scheduling remain challenging due to heterogeneous networks, dynamic workloads, and conflicting objectives.

This study provides a comprehensive review of resource management strategies in fog cloud environments, focusing on heuristic, metaheuristic, and learning-based approaches. Key aspects such as computation offloading, load balancing, energy optimization, and cost efficiency are analyzed, with attention to their strengths, limitations, and areas of application.

The study highlights key open challenges such as large-scale real-world validation, the incorporation of security and privacy mechanisms, cross-layer optimization, and the design of hybrid models that integrate metaheuristics with reinforcement learning. Through a structured taxonomy and critical analysis, it offers valuable guidance for advancing adaptive, secure, and energy-efficient resource management in future fog cloud ecosystems.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

Keywords:. *Fog Computing, Cloud Computing, Mist Computing, Fog–Cloud Ecosystem, Multi-layered Architecture, Edge Computing.*

1. Introduction

The rapid growth of the Internet of Things (IoT), smart cities, and latency-sensitive applications such as autonomous driving, augmented reality, healthcare monitoring, and smart grids have driven the need for computing infrastructures that can provide low-latency and energy-efficient services [1]. Traditional cloud computing, while offering elastic resources, often fails to meet stringent requirements owing to long transmission delays and centralized processing overhead. To address these limitations, fog computing has emerged as a complementary paradigm that extends the computation, storage, and networking capabilities closer to the network edge. The combination of fog, cloud, and mist computing forms a multi-layered ecosystem capable of supporting heterogeneous applications in dynamic and resource-constrained environments [6]. Within this ecosystem, resource allocation and task scheduling are recognized as critical challenges. The distributed nature of fog nodes, diverse QoS requirements, fluctuating workloads, and varying energy budgets complicate the efficient resource management. Numerous approaches have been proposed, ranging from heuristic algorithms such as round-robin scheduling and priority-based queuing to metaheuristic strategies such as Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Honey Badger Algorithm (HBA), Mexican Axolotl Optimization (MAO), and hybrid methods combining Harris Hawks Optimization (HHO) with PSO [4]. These algorithms demonstrate significant improvements in make span reduction, load balancing, energy efficiency, and cost minimization. However, they often face the challenges of convergence speed, scalability in ultra-large networks, and computational overhead [7].

Simultaneously, the rise of machine learning and reinforcement learning (RL) has opened new directions for adaptive and intelligent resource management. Deep Reinforcement Learning (DRL) and Deep Q-Learning (DQL) approaches have been employed to optimize task scheduling, resource allocation, and power consumption jointly, particularly in vehicular fog computing, fog radio access networks (F-RAN), and NOMA-enabled architectures [9]. These learning-driven methods offer dynamic adaptability to changing environments but also introduce new challenges such as training complexity, convergence delays, and the need for large-scale interaction data.

Several survey papers in the field have provided reviews on specialized aspects such as QoS-aware load balancing, matching theory for distributed offloading, and mutual satisfaction-based resource allocation in cloud–fog–mist environments [17]. These reviews highlight recurring limitations across the literature, including the lack of multi-objective optimization

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

frameworks that balance latency, energy, and cost simultaneously, absence of strong privacy and security guarantees, and limited real-world deployment studies [20].

This survey consolidates and extends the current state of research by reviewing 20 representative studies covering heuristic, metaheuristic, and learning-based approaches for resource allocation and scheduling in fog–cloud ecosystems. It categorizes existing solutions based on their objectives (latency minimization, energy efficiency, cost reduction, or fairness), methodologies (optimization-based, bio-inspired, or learning-driven), and deployment scenarios (vehicular fog, IoT networks, smart grids, healthcare, and F-RAN) [15]. The contributions of this survey are fourfold:

It offers a comprehensive taxonomy of resource allocation and scheduling strategies in fog–cloud computing, classifying them into optimization-based, metaheuristic, and learning-driven approaches [12]. It conducts a comparative assessment of these methods with respect to scalability, adaptability, and their effectiveness in ensuring Quality of Service (QoS). The discussion also brings forward cross-domain perspectives, showing how requirements and challenges vary across sectors such as vehicular networks, healthcare, and industrial IoT. In addition, the survey highlights future research opportunities, including the integration of optimization and learning techniques, the development of energy- and security-aware scheduling models, and the need for large-scale real-world validation. By consolidating these insights, this study delivers a holistic view of existing approaches while charting directions toward more robust, adaptive, and QoS-focused solutions in fog–cloud systems.

2. Related Work

Research on fog–cloud computing has produced a wide range of approaches for resource allocation and scheduling, each addressing specific challenges such as latency, energy efficiency, scalability, and Quality of Service (QoS). Several studies have focused on vehicular fog computing, where minimizing delay is crucial for safety-critical applications. For example, time-constrained scheduling schemes introduce metrics such as Perception-Reaction Time (PRT) and employ deep reinforcement learning (DRL) combined with Information-Centric Networking (ICN) to achieve significant latency reduction, though often at the cost of scalability owing to centralized control

In multi-layer fog–cloud systems, cooperative scheduling frameworks have been proposed to reduce bandwidth usage and balance workloads across end devices, fog nodes, and cloud servers. These methods improve scalability and efficiency but add complexity in coordination and remain underexplored in large-scale, dynamic environments

Fog Radio Access Networks (F-RAN) research has introduced scheduling schemes to reduce fronthaul load and outage probability, showing performance benefits but often limited to simplified single-node or static setups.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

Bio-inspired and metaheuristic algorithms also play a prominent role in fog–cloud task scheduling. Techniques such as Ant Colony Optimization (ACO), Honey Badger Algorithm (HBA), Mexican Axolotl Optimization (MAO), and hybrid strategies (e.g., PSO–HHO and IDLOA) have demonstrated improvements in processing time, throughput, and energy consumption. While these approaches highlight strong adaptability and convergence benefits, they frequently face the challenges of computational overhead, parameter sensitivity, and limited real-world validation

Learning-driven methods, especially reinforcement learning (RL) and multi-armed bandits (MAB), which have been applied to distributed fog environments to manage uncertainty and adapt to dynamic workloads. These models show promise for autonomous decision-making in heterogeneous systems, reducing dependence on centralized optimization. However, issues of slow convergence, training data demands, and scalability to dense IoT deployments remain open challenges. Other emerging directions involve matching theory for distributed offloading, Petri Net models for time- and cost-sensitive resource allocation, and mutual satisfaction frameworks that align user and provider goals. Although these approaches offer valuable perspectives, they commonly lack energy- and security-aware mechanisms, show limited progress in hybrid learning–optimization methods, and remain largely untested in large-scale real-world scenarios.

3. Methodology

The methodology illustrated in the diagram follows a taxonomy-driven approach to analysing resource allocation and scheduling in fog–cloud computing. At the top level, strategies are classified into two broad categories: optimization-based and learning-based methods.

Optimization-based approaches rely on deterministic or heuristic rules and metaheuristic search techniques. Heuristics such as round robin and priority queuing provide simplicity and low overhead, whereas metaheuristics such as PSO, ACO, HBA, MAO, and hybridized variants (e.g., HHO-PSO) aim to explore larger solution spaces and achieve better performance under dynamic workloads.

Learning-based approaches focus on adaptive decision-making through reinforcement learning methods (Q-Learning, DRL, Deep Q-Learning) and hybrid learning paradigms that combine machine learning with metaheuristics or leverage distributed schemes such as federated reinforcement learning. These methods emphasize adaptability, scalability, and QoS-awareness in highly heterogeneous environments.

The arrows in the diagram capture the flow from the methodological categories to their application domains, underscoring how each technique supports different real-world contexts. For instance, vehicular fog computing demands real-time and low-latency scheduling, F-RAN requires efficient bandwidth and task allocation, IoT and smart cities emphasize scalability and

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

healthcare systems prioritize reliability and security, while industrial IoT and edge AI call for robustness and efficiency.

By mapping techniques to applications, this methodology ensures a structured comparative analysis, highlighting both strengths and limitations. It also serves as a foundation for identifying research gaps such as the hybridization of optimization and learning methods, energy- and security-aware scheduling, and the importance of real-world large-scale validation.

Figure 1 illustrates the hierarchical flow of resource allocation and scheduling strategies in fog cloud computing, organized into two primary categories: optimization-based and learning-based approaches. Under optimization-based methods, heuristics (such as round robin and priority queuing) and metaheuristics (including PSO, ACO, HBA, MAO, and hybrid variants) are represented. Reinforcement learning techniques (Q-learning, DRL, Deep Q-learning) and hybrid learning approaches (integration of ML with metaheuristics or federated reinforcement learning) are highlighted.

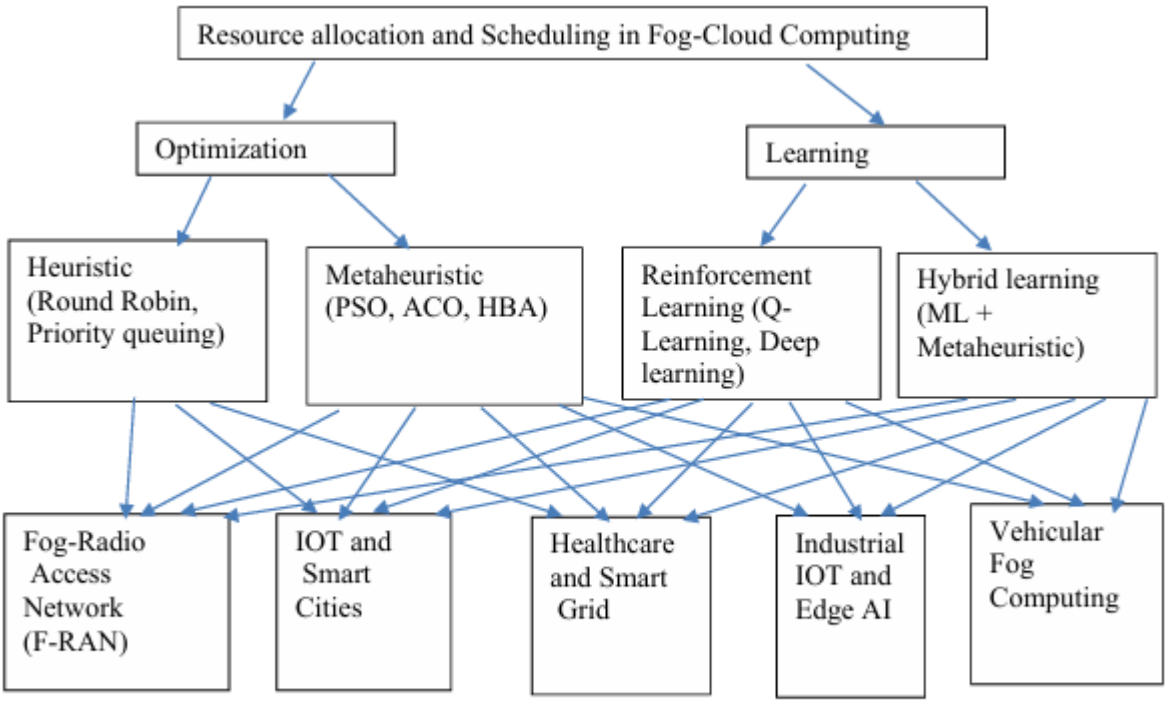

Figure 1: Methodological Framework of Resource Allocation and Scheduling Strategies in Fog Cloud Computing Across Application Domains

Both the categories and their sub-methods are linked through arrows to various application domains, showing their practical relevance. These domains include vehicular fog computing, fog radio access networks (F-RAN), IoT and smart cities, healthcare with smart grid systems, and industrial IoT with edge AI. The flow captures how methodological choices directly

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

support diverse real-world applications, providing a structured perspective on current research directions in fog–cloud environments.

Table 1 consolidates the major research contributions in fog–cloud resource management, summarizing their objectives, methodologies, advantages, limitations, and open research gaps. It covers a diverse range of strategies, including heuristic, metaheuristic, learning-based, and hybrid models, applied across vehicular networks, F-RAN, IoT, healthcare, and industrial IoT domains. Each entry highlights practical strengths such as scalability, energy efficiency and QoS improvements, while also pointing out constraints such as high computational overhead, scalability issues and limited real-world validation. By presenting both contributions and shortcomings, the table offers a clear comparative view that guides future research toward more adaptive, secure, and energy-efficient solutions.

Table 1: Comparative Summary of Resource Allocation and Scheduling Approaches in Fog Cloud Computing

ID	Title	Summary	Advantages	Disadvantages	Research gap
[1]	A Machine-Learning Based Time Constrained Resource Allocation Scheme for Vehicular Fog Computing	The paper addresses delay challenges in safety-critical vehicular applications within ITS by introducing Perception-Reaction Time (PRT) as a holistic delay metric. To minimize PRT, it proposes a vehicular fog computing framework integrating ICN with fog virtualization, supported by a DRL-based resource allocation strategy. Simulations show that this approach reduces PRT by	A new PRT metric accounts for both computation and communication delays. The use of ICN minimizes redundant data exchange, while DRL ensures adaptive resource allocation with faster convergence. The approach cuts PRT by about 70% and remains effective even under dense traffic conditions.	The approach relies on centralized DRL with a head vehicle or RSU, assumes uniform speed and constant queuing, and provides limited optimization for non-safety applications. Its strong dependence on RSUs also restricts scalability in ultra-dense vehicular networks.	Future work should evaluate the framework under 5G and LTE-V environments with realistic mobility and dynamic traffic models. Energy efficiency needs to be incorporated alongside an assessment of unresolved security and privacy concerns. Decentralized or federated RL can be explored to reduce reliance on central RSUs, while extending integration from fog to cloud for broader scalability. Additionally,

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

		nearly 70% compared to conventional methods			optimizing mixed workloads remains an open challenge.
[2]	A Novel Strategy to Achieve Bandwidth Cost Reduction and Load Balancing in a Cooperative Three-Layer Fog-Cloud Computing Environment	The paper proposes a three-tier cooperative fog–cloud framework designed to minimize bandwidth usage and balance workloads across fog nodes, cloud servers, and end devices. By combining task scheduling with data distribution, the approach enhances resource efficiency while maintaining service quality.	The framework improves scalability by coordinating fog and cloud resources, leading to better load distribution, lower bandwidth consumption, and more efficient service delivery in diverse computing environments.	The framework introduces added complexity from multi-layer coordination, and its scalability in large-scale networks is still uncertain. Moreover, practical issues like fluctuating latency and dynamic traffic conditions remain insufficiently explored.	The method applies cooperative scheduling, layered resource allocation, and bandwidth optimization. However, gaps remain in incorporating energy-efficient mechanisms, ensuring secure cross-layer collaboration, and validating performance at large-scale, real-world deployments.
[3]	An Efficient Scheduling Scheme for Fronthaul Load Reduction in Fog Radio Access Networks	This study introduces a scheduling scheme for fog radio access networks (F-RAN) that reduces fronthaul load while satisfying user throughput demands. The approach dynamically chooses the scheduler with minimal fronthaul usage, as demonstrated through simulation results.	The scheme decreases fronthaul load, minimizes outage probability, and ensures fairer user allocation. Compared to round-robin scheduling, it delivers lower latency and a better overall user experience.	The approach relies on simplified models such as a single F-AP and RRH, limiting its applicability in dense, multi-user, and dynamic channel settings. Additionally, energy efficiency remains unexplored.	The approach leverages F-APs, caching, and scheduling with optimization models. Future research should focus on large-scale validation, handling multi-user dynamics, and aligning with emerging 5G and beyond standards.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

[4]	Bandit Learning-Based Distributed Computation in Fog Computing Networks	This survey examines distributed computation offloading in fog networks through bandit learning, emphasizing multi-armed bandit (MAB) methods as an effective alternative to centralized optimization and game-theoretic models for handling uncertainty and dynamic conditions.	Bandit learning offers lightweight, distributed solutions that adapt to dynamic conditions and manage uncertainty effectively. By minimizing dependence on prior system knowledge, it enhances QoS and QoE in fog-enabled IoT applications..	Bandit algorithms encounter limitations in convergence speed, scalability across large networks, and managing the exploration–exploitation balance. They also face difficulties in handling task dependencies and heterogeneous fog resources.	The study employs contextual and non-contextual MAB algorithms for offloading, with gaps in DRL integration, energy efficiency, security, and large-scale real-world validation.
[5]	A Multi-Objective Fog Computing Task Scheduling Strategy Based on Ant Colony Algorithm	This paper proposes an enhanced ACO algorithm for fog task scheduling that integrates time-and-cost (TAC) metrics and a critical point factor to improve convergence. Simulations in iFogSim show reduced processing time, lower costs, and better load balancing than round-robin and standard ACO methods.	The algorithm achieves faster, more stable convergence by balancing time and cost, outperforming round-robin and standard ACO in efficiency and task distribution. Its validation with iFogSim confirms effectiveness in heterogeneous fog environments.	The study is limited to independent tasks, small-scale setups (≤10 nodes, 500 tasks), and a simplified cost model, reducing scalability and real-world relevance. Some algorithmic randomness also persists, affecting determinism.	The algorithm extends ACO with TAC-based multi-objective metrics, pheromone updates, and roulette wheel selection. However, it lacks support for task dependencies, detailed energy/cost models, large-scale validation, and security or fault tolerance. Future directions include hybrid metaheuristics and adaptive parameter tuning.
[6]	Energy and Resource	This paper introduces the	The method balances energy	HBA's reliance on random	The approach applies HBA with

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

	Aware Scheduling in Cloud-Fog Environment using Advanced Meta Heuristic Algorithm	Honey Badger Algorithm (HBA) for energy- and resource-aware scheduling in cloud–fog systems. Inspired by honey badger foraging, it balances exploration and exploitation for adaptive task scheduling. Simulations show HBA cuts energy use by 15.6%, boosts resource utilization by 18.2%, reduces task time by 17.4%, and raises throughput by 21.2% over PSO, GA, and IDOA.	and performance, adapts dynamically to workloads, and scales well in cloud–fog settings. It shows strong robustness and adaptability, outperforming traditional algorithms in simulations.	workflows can raise computational overhead, and its effectiveness in ultra-large, heterogeneous IoT networks remains unverified. It also depends on fine-tuning and may struggle in highly dynamic environments.	random workflow scheduling, validated through simulations. Key gaps include missing integration with real-world energy pricing, limited focus on security, and absence of large-scale deployment validation.
[7]	MAO – An Efficient Resource Utilization of Task Scheduling in Cloud-Fog Environment	This study presents the Mexican Axolotl Optimization (MAO) algorithm for cloud–fog task scheduling. Drawing on the axolotl's regenerative traits, it enables self-optimization and adaptability, enhancing resource use. Experiments show MAO cuts makespan by 28%, reduces response time by 36%, and boosts task success	MAO efficiently allocates resources, adapts to dynamic workloads, and improves responsiveness and reliability in cloud–fog systems. Its bio-inspired adaptability offers a novel solution to scheduling challenges.	The algorithm requires precise parameter tuning, faces uncertain scalability in large-scale real-world settings, and may incur higher computational complexity under heavy workloads.	The MAO approach uses bio-inspired optimization with random workflows for adaptability. However, it lacks large-scale real-world validation, standardized benchmarks, and exploration of hybrid optimization strategies for further enhancement.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

		by 31% over other methods.			
[8]	Cooperative Computing for Mobile Crowdsensing Design and Optimization	This paper presents a cooperative computing framework for mobile crowdsensing (MCS), enabling task offloading from source devices to nearby helpers with idle resources. By jointly optimizing offloading, communication, and computation, and solving the MINLP with DRESHA and alternating optimization, the approach reduces energy use and improves task completion compared to local execution.	The framework lowers energy consumption, boosts task completion, and reduces reliance on centralized edge or cloud systems by utilizing idle mobile devices to enhance overall efficiency.	The NP-hard nature of the problem and overhead from distributed implementation pose challenges, while real-world use in dynamic networks may struggle with stability and scalability.	The solution employs cooperative computing with OFDMA-based D2D links, matching theory, and alternating optimization. Key gaps include limited real-world validation, missing privacy and security integration, and the need for adaptive methods to manage high mobility and unpredictable conditions.
[9]	Fog-Based Resource Allocation Hybrid Approach Using Metaheuristic for Mobile Networks	This paper tackles fog resource allocation in mobile networks using a hybrid PSO–HHO approach. By distributing tasks across devices, fog nodes, and servers, it improves makespan, cost, and processing time. Simulations show it outperforms	The hybrid approach lowers makespan, cost, and processing time while boosting throughput and load balancing. It adapts effectively to dynamic fog settings and improves use of limited mobile resources.	The method incurs high overhead from hybrid optimization and constant load monitoring, with validation limited to small-scale testbeds, leaving scalability uncertain.	Developed on OMNeT++ and FOGNETsim++ with hybrid HHO–PSO, the approach still lacks large-scale real-world validation, energy-aware strategies, and security or privacy integration for mobile fog systems.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

		standalone PSO and HHO with better efficiency and load balancing.			
[10]	IDLOA – Prioritized Task Scheduling for Optimizing Resource Utilization in Cloud-Fog Environment	This paper presents the Improved Dingo Lion Optimization Algorithm (IDLOA), a hybrid bio-inspired method for cloud–fog task scheduling. By enhancing dynamic resource use, it lowers response time by 21%, energy consumption by 32%, and costs by 27% compared to existing approaches.	IDLOA adapts to heterogeneous cloud–fog environments, enhancing response time, energy efficiency, and cost savings. It balances exploration and exploitation to achieve robust task scheduling.	The hybrid design raises computational complexity, and its scalability in ultra-large deployments remains unproven. Its performance may also be sensitive to parameter settings, affecting robustness across diverse environments.	IDLOA merges two bio-inspired algorithms for prioritized scheduling. Gaps remain in real-world CFC validation, incorporation of QoS and deadline constraints, and exploring hybrids with ML-based methods.
[11]	MAO – An Efficient Resource Utilization of Task Scheduling in Cloud-Fog Environment	This study introduces the Mexican Axolotl Optimization (MAO) algorithm for cloud–fog task scheduling. Inspired by axolotl regeneration, it improves makespan, response time, and task success, achieving 28%, 36%, and 31% gains respectively over other methods.	MAO adapts to dynamic workloads, increases scheduling efficiency, and maximizes task success, with its bio-inspired design offering robustness against fluctuations.	The algorithm demands significant computation and careful parameter tuning, while its scalability in real-world settings remains uncertain.	MAO applies bio-inspired optimization with random workflows for adaptability. Key gaps include missing standardized benchmarks, limited study of hybrid strategies, and a lack of large-scale real-world validation.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

[12]	Optimizing Resource Allocation for Energy Efficiency in Fog-Cloud Computing Environments	This paper presents an energy-aware resource allocation framework for fog–cloud environments, optimizing task scheduling and VM allocation. By combining load balancing with energy-efficient scheduling, it reduces power consumption and shortens response times while maintaining service quality.	The framework cuts energy use, balances workloads, and speeds up task responses, promoting greener computing in large-scale distributed systems.	The model grows complex in highly dynamic settings, and its scalability to large heterogeneous or multi-tenant networks remains unverified.	The framework combines energy-aware scheduling with VM migration, but gaps remain in real-world validation, use of AI-based predictive energy models, and testing under varied QoS constraints.
[13]	Reinforcement Learning-Based Resource Management Model for Fog Radio Access Network Architectures in 5G	This study introduces a reinforcement learning (RL) approach for dynamic resource management in fog radio access networks (F-RAN) for 5G. Using Q-learning, the model autonomously allocates computational resources to minimize latency and optimize resource utilization. Simulations in smart farming scenarios show a reduction in data transfer to the cloud by up to 90%, highlighting the	It enables autonomous, adaptive resource management that lowers latency, improves utilization in dynamic 5G IoT settings, reduces cloud dependence, and enhances services in underserved areas.	Q-learning converges slowly in complex settings, demands extensive interaction data for training, and may underperform with highly heterogeneous workloads.	The approach applies Q-learning with 5G K-SimNet and OpenAI Gym. Gaps remain in validating scalability for dense urban IoT, integrating deep RL, and incorporating energy efficiency into RL-based resource allocation.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

		potential for IoT-heavy applications			
[14]	Research on Cloud Computing Resource Allocation Based on Particle Swarm Optimization Algorithm	This paper applies Particle Swarm Optimization (PSO) to cloud resource allocation, dividing tasks into subtasks and dynamically assigning resources to cut costs and execution time. C++ simulations demonstrate improved efficiency in large-scale data centers.	It improves allocation efficiency, shortens task completion time, and adapts to dynamic workloads, delivering globally optimal scheduling with scalability across data centers.	PSO risks premature convergence, incurs higher overhead at scale, and has limited capability to manage multiple objectives such as energy and security together	The approach uses PSO within cloud scheduling frameworks. Gaps remain in extending to hybrid cloud–fog models, integrating with other metaheuristics, and validating in real-world multi-tenant environments.
[15]	Resource Allocation Strategy in Fog Computing Based on Priced Timed Petri Nets	This paper presents a PTPN-based strategy for fog resource allocation, modeling both time and cost constraints to enable dynamic, efficient service provisioning. It validates task allocation and resource use in real-time settings.	It models time- and cost-sensitive applications accurately, enhances real-time allocation decisions, and supports concurrent operations in fog environments.	PTPN models involve high computational complexity and demand precise parameter settings, limiting scalability and leaving their applicability in large, heterogeneous fog systems uncertain.	The approach uses PTPN formal modeling for fog scheduling. Key gaps include AI-based optimization integration, support for diverse service demands, and large-scale real-world validation.
[16]	A Comprehensive Review of QoS Aware Load Balancing Techniques in Generic & Specific Fog Deployment Scenarios	This survey reviews QoS-aware load balancing in fog computing across domains like smart grids, EVs, and healthcare. It evaluates methods by metrics such as response time, latency, energy, and cost, and classifies	It offers a taxonomy of load balancing strategies, outlines their strengths across domains, and identifies key QoS metrics, serving as a guide for selecting appropriate	Most methods struggle to balance multiple QoS trade-offs like latency, cost, and energy, and are largely limited to simulation-based validation without large-scale	It reviews heuristic (Round Robin, Priority Queuing), metaheuristic (PSO, ACO, GA), and learning-based (RL) methods. Gaps remain in multi-objective optimization, real-world scalability, and integrating

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

		them into heuristic, metaheuristic, and ML-based approaches.	techniques in different scenarios.	deployment testing.	security and privacy into QoS load balancing
[17]	A Survey on Matching Theory for Distributed Computation Offloading in IoT-Fog-Cloud Systems – Perspectives and Open Issues	This survey examines matching theory for distributed computation offloading in IoT–Fog–Cloud systems, highlighting it as a lightweight alternative to centralized optimization and game-theoretic methods for handling heterogeneity, scalability, and distributed decisions.	Matching-based models lower computational complexity, enable distributed resource allocation, and ensure stable task–resource pairing, making them well-suited for heterogeneous fog and IoT settings.	Matching theory emphasizes stability over global optimality, potentially limiting efficiency, and depends on clear preference structures that may be impractical in dynamic environments.	It uses one-to-one, one-to-many, and many-to-many matching models with distributed algorithms. Gaps remain in integrating AI/ML for adaptability, incorporating privacy and security, and validating at scale in next-generation fog–IoT systems.
[18]	Dynamic Energy Efficient Resource Allocation Strategy for Load Balancing in Fog Environment	This paper introduces the Dynamic Energy-Efficient Resource Allocation (DEER) strategy for fog computing. Featuring modules like Task Manager, Resource Scheduler, Engine, and Power Manager, it dynamically allocates resources based on utilization, cutting costs and energy use compared to DRAM.	It lowers energy use by ~8.67% and computational costs by ~16.77%, enhancing task execution efficiency with energy-aware scheduling and resource management.	Its complexity grows in large dynamic settings and depends on predefined task and resource costs, which may not accurately reflect heterogeneous IoT environments.	Built on fog computing with energy-aware scheduling and on/off power mechanisms, the approach still lacks real-world validation, scalability for ultra-large networks, and integration with AI-based predictive models.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

[19]	MSRM-IoT – A Reliable Resource Management for Cloud, Fog, and Mist-Assisted IoT Networks	This paper presents MSRM-IoT, a mutual satisfaction-based framework for cloud–fog–mist IoT networks. By combining user and provider preferences for latency, cost, and reliability into a unified satisfaction function, and using a layered architecture (consumer, mist, edge, fog, cloud), it optimizes QoS while balancing trade-offs.	It balances user satisfaction (latency, cost) with provider goals (utilization, energy efficiency), delivering higher QoS and fairness than traditional allocation methods.	Managing multiple satisfaction functions adds computational overhead, and reliance on detailed preference and resource data limits scalability and real-time adaptability in dynamic IoT settings.	It applies a layered fog–cloud–mist design with a mutual satisfaction function and optimization algorithms. Gaps remain in large-scale real-time validation, stronger privacy/security integration, and adaptation to highly dynamic workloads.
[20]	Task Offloading in NOMA-Based Fog Computing Networks – A Deep Q-Learning Approach	This study investigates fog task offloading using NOMA with Deep Q-Learning, jointly optimizing scheduling, resource, and power allocation to balance energy and delay. Simulations show notable cost reductions over baseline strategies.	It enhances spectrum efficiency with NOMA, cuts offloading delays, and achieves a better energy–delay trade-off, using reinforcement learning to adapt to dynamic wireless conditions.	DQL demands heavy training and may struggle with convergence, while NOMA adds interference management challenges, with performance potentially degrading in highly heterogeneous networks.	The approach integrates fog computing with NOMA for spectrum efficiency and DQL for adaptive resource management. Gaps remain in scaling to dense IoT networks, ensuring robustness in variable wireless conditions, and exploring hybrid offloading strategies.

4. Related Outcomes

The proposed REALM-FC (Real-time Energy-, Adaptive, and Latency-aware Multi-objective Scheduling in Fog–Cloud Computing) algorithm directly addresses the challenges of balancing

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

delay, energy efficiency, cost, and QoS in heterogeneous fog–cloud environments. By combining real-time monitoring with predictive estimates of computation, queueing, and network delays, it adaptively allocates tasks to fog or cloud nodes while minimizing overhead. Its multi-objective score function ensures that deadlines, utilization, and energy budgets are respected, making it suitable for latency-critical and resource-constrained applications. Positioned at the intersection of fog computing, cloud-edge orchestration, and resource optimization, REALM-FC contributes to advancing research in IoT, vehicular networks, and next-generation distributed systems.

Algorithm 1 : Real-time Energy-, Adaptive, and Latency-aware Multi-objective Scheduling in Fog Cloud Computing

Step 1 **Processing Time Estimation:** Assess the time required for a task to execute on a fog or cloud node *j*, based on the task's workload volume and the node's available processing power.

$$T_{i,j}^{comp} = \frac{\boldsymbol{d_i}}{\mu_j}$$

Step 2 Queuing delay estimation: Approximate the waiting time a task experiences in the buffer of a fog or cloud node before it begins execution.

$$T_{i,j}^{queue} = \frac{q_j}{\mu_j} \qquad \text{with } Q_j \leftarrow Q_j + d_i \; if \; t_i assigned \; to \; j$$

Step 3 Network Delay and Data Transfer Evaluation:- For fog nodes, account for the local transmission delay. In the case of cloud nodes, they includes both the data transfer time over the network and additional round-trip communication delay.

$$T_{i,j}^{net} = L_{u \leftarrow j} + \frac{\boldsymbol{S_i}}{B_{u \leftarrow j}}$$

For cloud

$$T_{i,c}^{net} = L^{WAN} + \frac{\boldsymbol{S_i}}{B_{u \leftarrow c}}$$

Step 4 End-to-End Latency Prediction:- Derive the overall response time by aggregating the computation duration, queueing delay, and communication/transmission time.

$$\widehat{\boldsymbol{L_{ij}}} = T_{i,j}^{net} + T_{i,j}^{queue} + T_{i,j}^{comp}$$

Step 5 Energy Consumption Estimation:- Assess the energy required for both task execution and data transmission, considering the processing cycles consumed and the associated communication overhead.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

$$\widehat{E_{i,j}} = \alpha_j d_i + \beta_{u \leftarrow j} \beta_i$$

$$\beta \text{ is energy per transmitted bit}$$

Step 6 Feasibility Verification: - Confirm whether the task satisfies the defined latency and energy constraints. If the conditions are satisfied the task is scheduled; otherwise, it is either reassigned to another node or discarded.

$$\widehat{L_{i,j}} \leq D_i \quad and \quad U_j = \frac{Q_j}{\mu_j H} \leq U^{max}$$

$$U_j \text{ is Utilization estmate over horizon } H$$

The algorithm 1 illustrates that the Real-time Energy-, Adaptive, and Latency-aware Multi-objective Scheduling algorithm in fog–cloud computing enhances task scheduling by minimizing latency, reducing energy consumption, balancing workloads, and ensuring feasibility under time, energy, and resource constraints.

Research Areas

1. **Fog–Cloud Resource Management**
   Central to this survey is the coordination of computing, storage, and network resources across fog nodes and cloud datacentres. This area examines the strategies for balancing workload distribution, optimizing task placement, and improving interoperability in heterogeneous and geographically distributed environments.

2. **Metaheuristic Optimization**
   Algorithms such as PSO, ACO, HBA, MAO, and hybrid methods, are widely applied for scheduling tasks under uncertain and dynamic conditions. This research area focuses on efficiency finding near-optimal solutions where exact methods are computationally infeasible.

3. **Learning-Based Approaches**
   It covers reinforcement learning (Q-learning, DRL, deep Q-learning), multi-armed bandits, and hybrid ML + metaheuristic strategies. Emphasis is placed on adaptability, scalability, and handling dynamic task arrivals and uncertain network states in real time.

4. **Quality of Service (QoS) and Quality of Experience (QoE)**
   It ensures that applications in vehicular networks, healthcare, industrial IoT, and smart cities meet requirements such as latency, throughput, fairness, and reliability.
   QoS-aware approaches model user satisfaction and system performance simultaneously, making this a cross-cutting research area.

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

5. **Energy-Efficient and Sustainable Computing**
   A growing area of research has focused on reducing energy consumption and carbon footprint in fog–cloud environments. Techniques such as energy-aware scheduling, green task migration, and dynamic power management have been studied for sustainable deployments.

6. **Security- and Privacy-Aware Scheduling**
   It addresses threats in distributed fog–cloud networks, such as data leakage, insecure offloading, and malicious nodes. This emerging research area integrates trust-aware scheduling and secure task allocation alongside performance optimization.

7. **Application-Specific Resource Allocation**
   Domains such as vehicular fog computing the healthcare IoT, smart grids, and industrial automation impose unique constraints. Research has explored tailoring allocation strategies for domain-specific latency, reliability, and scalability.

5. Conclusion

This survey systematically reviewed resource allocation and scheduling strategies in fog–cloud computing and classified them into metaheuristic, learning-driven, and QoS-aware approaches. Metaheuristic methods demonstrate efficiency in solving large-scale optimization problems but often face scalability and real-time adaptability issues. Learning-based techniques, particularly reinforcement learning and hybrid models, show strong adaptability to dynamic and uncertain environments but require substantial training data and careful convergence handling. QoS-aware strategies emphasize latency, energy, and user satisfaction. However most remain validated only in simulation rather than real-world deployments. The comparative analysis highlights that no single approach is universally optimal; instead, hybrid frameworks that combine optimization, metaheuristics, and machine learning hold the greatest promise. Future research directions include integrating energy- and security-aware scheduling, developing federated and distributed learning solutions, and conducting large-scale real-world validations to bridge the gap between theory and deployment. By synthesizing insights across these domains, this survey provides a foundation for advancing robust, adaptive, and QoS-driven resource management in next-generation fog–cloud systems.

References

[1] X. Chen, S. Leng, K. Zhang, and K. Xiong, “A Machine-Learning Based Time Constrained Resource Allocation Scheme for Vehicular Fog Computing,” *China Communications*, vol. 16, no. 11, pp. 1–14, 2019.

[2] M. M. S. Maswood, M. R. Rahman, A. G. Alharbi, and D. Medhi, “A Novel Strategy to Achieve Bandwidth Cost Reduction and Load Balancing in a Cooperative Three-

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

Layer Fog-Cloud Computing Environment," *IEEE Access*, vol. 8, pp. 113737–113750, 2020,

[3] B. S. Bhandari, H. P. Zhao, and H. Kim, "An efficient scheduling scheme for fronthaul load reduction in fog radio access networks," *China Communications*, vol. 16, no. 11, pp. 146–153, Nov. 2019.

[4] H. Tran-Dang, K.-H. Kwon, and D.-S. Kim, "Bandit Learning-Based Distributed Computation in Fog Computing Networks: A Survey," *IEEE Access*, vol. 11, pp. 1-17, 2023.

[5] J. Gu, J. Mo, B. Li, Y. Zhang, and W. Wang, "A multi-objective fog computing task scheduling strategy based on ant colony algorithm," in *2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Dalian, China, Sept. 24–26, 2021, pp. 12–16,

[6] C. R. Bennett, L. Parvathaneni, R. R. Oduru, N. B. Riquelme, D. McLaughlin, and S. K. Medishetti, "Energy and Resource Aware Scheduling in Cloud-Fog Environment using Advanced Meta Heuristic Algorithm," in *Proc. 2024 2nd Int. Conf. Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 2024, pp. 1–6,

[7] K. Anusha, P. Archana, G. R. Karri, S. K. Medishetti, and K. P., "MAO: An Efficient Resource Utilization of Task Scheduling in Cloud Fog Environment," in *Proc. 2024 Int. Conf. Wireless Communications Signal Processing and Networking (WiSPNET)*, 2024, pp. 1–6,

[8] X. Xie, T. Bai, W. Guo, Z. Wang, and A. Nallanathan, "Cooperative Computing for Mobile Crowdsensing: Design and Optimization," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 6437–6452, May 2024

[9] Z. H. Ebrahim and M. E. Manaa, "Fog-based Resource Allocation Hybrid Approach using Metaheuristic for Mobile Networks," in *Proc. 2023 6th Int. Conf. Engineering Technology and its Applications (IICETA)*, Babylon, Iraq, 2023, pp. 1–6

[10] S. K. Medishetti, R. K. Donthi, C. M. Reddy, G. R. Karri, T. S. Vardhan, and K. V. Kumar, "IDLOA: Prioritized Task Scheduling for Optimizing Resource Utilization in Cloud-Fog Environment," in *Proc. 2024 IEEE 13th Int. Conf. Communication Systems and Network Technologies (CSNT)*, 2024, pp. 1–6

[11] K. Anusha, G. R. Karri, P. Archana, S. K. Medishetti, and K. P., "MAO: An Efficient Resource Utilization of Task Scheduling in Cloud Fog Environment," in *Proc. 2024 Int. Conf. Wireless Communications Signal Processing and Networking (WiSPNET)*, 2024, pp. 1–6

[12] M. Kaur, R. Aron, and S. Seth, "Optimizing Resource Allocation for Energy Efficiency in Fog Cloud Computing Environments," in *Proc. 2024 IEEE 13th Int. Conf. Communication Systems and Network Technologies (CSNT)*, 2024, pp. 1–6,

[13] N. N. Khumalo, L. Mfupe, and O. O. Oyerinde, "Reinforcement Learning-Based Resource Management Model for Fog Radio Access Network Architectures in 5G," *IEEE Access*, vol. 9, pp. 13318–13333, Jan. 2021,

[14] H. Fu, Y. Fan, X. Pan, Y. Tian, L. Shi, C. Xiao, and Q. Huang, "Research on Cloud Computing Resource Allocation Based on Particle Swarm Optimization Algorithm," in *Proc. 2021 IEEE Int. Conf. Advances in Electrical Engineering and Computer Applications (AEECA)*, 2021, pp. 1–6,

Power System Technology
ISSN:1000-3673
Received: 16-09-2025 Revised: 05-10-2025 Accepted: 26-11-2025

[15] L. Ni, J. Zhang, C. Jiang, C. Yan, and K. Yu, "Resource Allocation Strategy in Fog Computing Based on Priced Timed Petri Nets," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1216–1228, Oct. 2017,

[16] G. Goel and A. K. Chaturvedi, "A Comprehensive Review of QoS Aware Load Balancing Techniques in Generic & Specific Fog Deployment Scenarios," in *Proc. 2023 Int. Conf. Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 2023,

[17] H. Tran-Dang and D.-S. Kim, "A Survey on Matching Theory for Distributed Computation Offloading in IoT-Fog-Cloud Systems: Perspectives and Open Issues," *IEEE Access*, vol. 10, pp. 118353–118369, 2022,

[18] A. U. Rehman, Z. Ahmad, A. I. Jehangiri, M. A. Alaanzy, M. Othman, A. I. Umar, and J. Ahmad, "Dynamic Energy Efficient Resource Allocation Strategy for Load Balancing in Fog Environment," *IEEE Access*, vol. 8, pp. 199,829-199,839, 2020,

[19] A. S. M. Sanwar Hosen, P. Kumar, and G. H. Cho, "MSRM-IoT: A Reliable Resource Management for Cloud, Fog, and Mist-Assisted IoT Networks," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2527-2537, Feb. 15, 2022,

[20] K. Wang, Y. Zhou, Y. Yang, X. Yuan, and X. Luo, "Task Offloading in NOMA-Based Fog Computing Networks: A Deep Q-Learning Approach," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2019.