A Review Load balancing algorithms in Fog Computing

Roa'a Mohammed Mahdi*, Hassan Jaleel Hassan and Ghaidaa Muttasher Abdulsaheb
Computer Engineering Department, University of Technology, Baghdad, Iraq

Abstract. With the rapid advance of the Internet of Things (IoT), technology has entered a new era. It is changing the way smart devices relate to such fields as healthcare, smart cities, and transport. However, such rapid expansion also challenges data processing, latency, and QoS. This paper aims to consider fog computing as a key solution for addressing these problems, with a special emphasis on the function of load balancing to improve the quality of service in IoT environments. In addition, we study the relationship between IoT devices and fog computing, highlighting why the latter acts as an intermediate layer that can not only reduce delays but also achieve efficient data processing by moving the computational resources closer to where they are needed. Its essence is to analyze various load balancing algorithms and their impact in fog computing environments on the performance of IoT applications. Static and dynamic load balancing strategies and algorithms have been tested in terms of their impact on throughput, energy efficiency, and overall system reliability. Ultimately, dynamic load balancing methods of this sort are better than static ones for managing load in fog computing scenarios since they are sensitive to changing workloads and changes in the system. The paper also discusses the state of the art of load balancing solutions, such as secure and sustainable techniques for Edge Data Centers (EDCs). It manages the allocation of resources for scheduling. We aim to provide a general overview of important recent developments in the literature while also pointing out limitation where improvements might be made. To this end, we set out to better understand and describe load balancing in fog computing and its importance for improving QoS. We thus hope that a better understanding of load balancing technologies can lead us towards more resilient and secure systems.

1 Introduction

The Internet of Things (IoT) has emerged as a transformative technology, reshaping various aspects of our lives and industries. IoT consists of a network of interconnected devices that talk to each other, each carrying sensors, actuators, and processors working together for specific purposes. In smart homes, health care systems, transportation, agriculture, and countless other uses, the number of these devices exceeds people [1][2][3]. The IoT’s potential for generating economic value is huge and fast-growing; it could yield $5.5 to $12.6 trillion in value to the world by 2030 [4], and this comprises not only producers of products and services but also the value that consumers and users obtain from them. However, IoT’s potential was hindered internally by shortcomings in storage capacity, processing capabilities, and privacy policies necessary for linking with cloud services [5][6]. In this case, fog computing presents a solution to the current constraints of the cloud in IoT settings. Fog computing is a transitional layer between cloud-brained self-driving cars and IoT devices, destinations bringing crucial resources nearby and decreasing the lag time, thus improving privacy and security [7][8][9]. Fog computing's ability to combine and analyze the significant amounts of data that these machines continuously generate highlights the over-dependence and bottleneck issues of connecting directly via IoT devices. It offers a cost-efficient means to manage data needing more economical analysis [10]. Fog technology was designed specifically for IoT, with features like low latency and geo-distributed systems or location awareness that are lacking in traditional cloud infrastructure [11][12]. Load balancing is a technology to uniformly distribute workloads across many different computers or links in a network so that they are all used equally often. This optimizes resource consumption, maximizes throughput, minimizes response time, and prevents overload. To achieve load balancing on a network, key parameters include throughput, bandwidth, Capacity and Performance (CP) consumption, number of drop packets, packet delay rate (PDR), and end-to-end delay [13]. Fog computing's introduction to the IoT world hasn't been easy, and there have been many difficulties along the way. What is needed are complex designs of cloud-fog architectures or resource allocation; further, these have to support gigantic distributed control systems controlling everything from sails to light bulbs [9][14]. This is why fog computing must be incorporated into IoT designs in order to overcome these problems. It provides a more efficient means of collecting, processing and integrating data for distribution [15][15][16]. In light of this challenge, this paper focuses on surveying the IoT applications and services that use fog computing, concentrating particularly on load balancing. Load balancing holds great importance in fog computing with respect to resource

* Corresponding author: ce.21.04@grad.utechnology.edu.iq
distribution, functional performance, and Quality of Service (QoS). A variety of load balancing mechanisms and techniques within the domain of fog computing and their implications for QoS in IoT applications are examined in this paper. In order to examine the existing literature, case studies, and practical applications and add to the field, this paper will guide readers into deeper corridors of understanding about the apotheosis of IoT up through fog computing; it will provide guidelines on how load balancing takes shape in this network[17]. In which the ultimate goal is to catalyze future work that combines technology and user experience of performance. Yet the IoT has created a pressing need to offer end-to-end services while minimizing network latency. The field of fog computing has developed different strategies for load balancing or the provision of quality of service (QoS) according to its unique problems and environment [18][19].

Researchers have categorized load balancing into static and dynamic types. Studies show that static techniques have shortcomings with varying loads [20]. Security and sustainability of the load-balancing technique have been proposed in various ways so far. Specifically, authenticating Edge Data Centers (EDCs) and allocating tasks efficiently to underutilized EDCs [21]. Many studies have been conducted to evaluate the effectiveness of different allocation techniques. Static, random, and proportional allocations are among the approaches to be compared. Based on those studies, it is widely held that proportional allocation is superior to the others. In addition to these studies, researchers have looked at other advanced techniques to solve this problem: genetic algorithms or Q-learning for dynamic resource allocation in fog-based medical environments [22][23][24]. Besides, breadth first search (BFS) techniques in designing load balancing systems were also used for this purpose [25] [21][26]. Understanding traditional cloud computing load balancing algorithms and applying them to the dynamic and heterogeneous nature of fog networks has long been a primary concern of researchers [27]. With the change from static to dynamic load balancing methods, many researchers have been exploring solutions in this area, and dynamic methods have been proven to be more effective of the two technologies because they can better adapt to variations in user behavior and system conditions [28]. A variety of techniques, including service migration, load optimization, and load balancing in fog computing, have all been reviewed by comprehensive surveys [29][30]. This underscores the breadth and dynamism of the research landscape. The application of Software-Defined Networking (SDN) in load balancing has shown promising results in improving delay reduction [31], [32], [33], while techniques based on dynamic graph partitioning have been developed to mitigate issues like node migration [29][34]. Studies have also focused on specific applications, such as Load balancing for real-time mobile face recognition in cloud-fog networks [35], and distributed Load balancing in smart city infrastructures[36]. Energy efficiency has been another critical focus, with research into dynamic, energy-efficient resource allocation strategies for Load balancing in fog environments [37][38]. Some researchers have also explored new methods of combining algorithms designed for application in healthcare with conventional real-time caching and scheduling didactics [39]. However, we still have many holes to fill in our understanding of how to balance network load based on various complex demands arising from the deployment of fog systems. Yet, future efforts must take up these challenges and develop load control algorithms that can adapt to the quickly changing needs of fog computing environments.

2 Literature Review

Numerous researchers have made substantial contributions to the challenges of load balancing in the developing field of fog computing. The literature on load balance in Fog is reviewed in the following section. Anees Ur Rehman (2020) [40] suggested that Dynamic Energy Efficient Resource Allocation (DEER) is a strategy for maintaining load balance and conserving energy resources in fog-computing environments. This strategy groups tasks and assigns them to resources based on their level of utilization. As a result, the load is balanced. It has a Resource Power Manager, which, during task execution, regulates resource power states in order to save energy. Those effects are equivalent to a reduced energy consumption of 8.67% and computation costs of 16.77% as compared to the DRAM method used currently. The core innovations are its load-balancing technology and power-optimization mechanism.

In the same year, Fatma M. Talaat (2020) [41] introduced a load balancing and optimization strategy (LBOS) for fog computing using reinforcement learning and genetic algorithms. LBOS dynamically allocates resources to balance load across servers, enhancing performance. It is designed for healthcare Internet of Things (IoT)-fog systems with three layers: IoT devices, fog nodes, and the cloud. Experiments show LBOS reduces allocation costs and response times, achieving 85.71% load balancing - better than existing algorithms. Key innovations are the hybrid machine learning resource optimization and application to healthcare IoT-fog architectures. LBOS efficiently utilizes resources and ensures continuous service availability.

Mirza M. Maswood (2020) [42] developed an optimization model for a 3-layer fog-cloud computing architecture to minimize resource costs and reduce delays for real-time IoT applications. The model balances loads across network bandwidth and server CPU capacity. Simulations show the framework effectively utilizes the cooperative fog-cloud environment to reduce bandwidth costs and balance loads by assigning priority weights to the composite objective function. It analyzes tradeoffs between minimizing costs versus improving load balancing. Key innovations are the multi-objective optimization and analysis of a tiered fog-cloud IoT infrastructure for cost and performance. The model allows flexible resource allocation based on application requirements.

Advancing to 2021, D. Baburao et al. (2021)[43] suggested an Enhanced Dynamic Resource Allocation Method (EDRAM) based on particle swarm optimization to address load balancing challenges in fog computing architectures serving many IoT devices. Fog computing is deployed between IoT devices and the cloud to reduce latency. EDRAM removes inactive services from memory and efficiently allocates resources to balance loads across fog nodes. This reduces...
task waiting times, latency, network bandwidth consumption, and improves Quality of Experience. The key innovation is using particle swarm optimization for dynamic resource allocation in fog computing to better manage large numbers of IoT devices.

Ahmad Raza Hameed (2021) [44] presented a cluster-enabled capacity-based load balancing for vehicular fog computing, where moving vehicles act as fog nodes processing IoT jobs. Dynamic clusters are formed based on vehicle positions, speed, and direction. Departure times are predicted to manage resources. Load balancing occurs intra- and inter-cluster. Simulations in NS2 show the approach balances energy consumption, reduces delay, and improves utilization versus other methods. Key innovations are dynamic clustering and predictive load balancing for vehicular fogs. The scheme efficiently leverages vehicle mobility and compute resources for IoT processing.

In 2022, Prabhdeep Singh [45] proposed a framework with user, cloud, and fog subsystems to minimize resource costs and balance loads in fog computing architectures for real-time Internet of Things applications. A fog cluster-based load balancing approach with a refresh period is introduced to optimize fog resource utilization. Simulations show this technique decreases energy consumption, VM migrations, and host shutdowns versus other methods. Key innovations are the tiered framework and fog cluster load balancing approach. By efficiently leveraging fog resources, the method reduces costs and balances dynamic loads for time-critical IoT services.

B. Kanbar and K. Faraj (2022) [46] developed a 5-step RADISH model for load balancing in cloud computing architectures serving Internet of Things devices. It classifies tasks using a neural network, schedules them with a moth flame optimization algorithm, balances loads with a potential field clustering approach, and allocates resources with a Hopcroft-Karp algorithm. Simulations in CloudSim show RADISH reduces latency, completion times, energy use, and SLA violations versus other methods. Key innovations are the hybrid machine learning techniques for optimized task scheduling, VM clustering for load distribution, and state-aware allocation. By efficiently leveraging cloud resources, RADISH improves QoS and SLAs for time-critical IoT services.

Fatma M. Talaat (2022), in a separate contribution [47], introduced an Effective Dynamic Load Balancing technique (EDLB) for fog computing architectures, composed of a fog resource monitor, CNN-based classifier, and optimized dynamic scheduler. EDLB monitors server resources, classifies suitability using convolutional neural networks, and schedules processes to balance loads. Simulations for healthcare applications show EDLB reduces response times and improves resource utilization versus other methods. Key innovations are the neural network-based filtering of fog nodes and dynamic optimization of task scheduling. By efficiently distributing processes, EDLB ensures continuous service availability and high performance for real-time systems.

Simar Preet Singh (2022) [48] proposed a Fuzzy Golden Eagle Load Balancing (FGELB) strategy for fog computing architectures serving many Internet of Things devices. FGELB uses a fuzzy algorithm to prioritize tasks based on deadlines, sizes, and predefined priorities. A Golden Eagle Optimization ranks resources for scheduling to ensure tasks execute on suitable nodes. A power engine manages resource power states. Simulations show FGELB reduces energy use, failures, costs, overheads, and waiting times versus existing methods. Key innovations are the hybrid fuzzy logic and bio-inspired optimization for task prioritization and scheduling. By efficiently leveraging fog resources, FGELB improves quality-of-service for real-time IoT applications.

Yu. Dongmin (2022) [49] propose a 3-tier architecture for load balancing in fog computing systems, consisting of a cloud layer, fog layer, and consumer layer. A real-time VM migration algorithm is introduced to balance loads across fog nodes and microgrids supplying consumers. Compared to closest data center and dynamic reconfiguration methods, simulations show the algorithm improves cost and response time performance by 18% and 11% respectively. Key innovations are the multi-tier architecture and dynamic migration approach for optimized, real-time resource allocation in fog networks. By efficiently distributing VMs, the load balancing scheme reduces costs and latency for time-critical services.

Moving to 2023, Ebrahim and A. Hafid (2023) [50] proposed an ELECTRE multi-criteria decision analysis approach for load balancing in fog computing architectures serving Internet of Things devices. It distributes workloads across limited fog resources and service replicas based on multiple objectives including compute and network loads. Simulations in realistic unbalanced topologies show it improves overall system performance by 67% versus common methods like random, round-robin, nearest, and fastest service selection. Key innovations are leveraging multiple objectives and ELECTRE methodology for optimized real-time fog resource allocation. By efficiently assigning tasks, the scheme enhances quality-of-service for delay-sensitive distributed applications.

Fatemeh R. Shahidani (2023) [51] developed a reinforcement learning-based scheduling algorithm for fog computing architectures serving many Internet of Things devices. Fog nodes between devices and the cloud aim to reduce data center loads and request latencies. The algorithm addresses key challenges of request scheduling, load balancing, and energy efficiency. Experiments show it improves load distribution and diminishes response times versus existing methods by optimally utilizing fog resources. Key innovations are leveraging reinforcement learning for dynamic resource allocation in fog networks. By efficiently assigning requests, the algorithm enhances quality-of-service and reliability for real-time IoT applications.

Monica R. Mundada et al. (2023) [52] explored a Mutated Leader load balancing algorithm (MLA) for fog computing architectures. MLA leverages Deep Residual Networks to predict VM resources and uses a leader-based approach to optimally allocate user tasks, balancing loads across fog cluster nodes. When VMs become overloaded, jobs are redistributed to underloaded VMs. Experiments show MLA minimizes execution time, cost, and load percentage versus other methods. Key innovations are the integration of deep learning for resource prediction and bio-inspired optimization
for dynamic task assignment in fog networks. By efficiently distributing jobs, MLA enhances performance and utilization for fog computing systems.

Finally, Ismail Zahraddeen Yakubu & M. Murali (2023) [53] present a layer fit algorithm to distribute tasks between fog and cloud computing resources based on priority, preventing oversaturation. A Modified Harris-Hawks Optimization (MHHO) then assigns resources within layers. Simulations show the framework reduces makespan, costs, energy usage, and improves resource utilization versus Harris-Hawks Optimization and other methods. Key innovations are the dual optimization approach leveraging task prioritization and bio-inspired meta-heuristics for dynamic fog-cloud resource allocation. By efficiently load balancing across tiers, the scheme enhances quality-of-service for real-time Internet of Things applications.

To have a comprehensive understanding of the trend, limitations and other factors of the reviewed previous works, Table 1 is presented.

The development of load balancing in fog computing has progressed from simple optimization models to more complex, comprehensive strategies, as described in the literature. Early works focused on minimizing costs and improving efficiency through dynamic resource allocation. As research progressed, the incorporation of machine learning, cluster-based approaches, and algorithmic advancements marked a significant leap. The latest studies emphasize multi-objective optimization and sophisticated algorithms, reflecting a nuanced understanding and an increasingly targeted approach to the challenges in fog computing.

**Table 1. Previous work summarization.**

<table>
<thead>
<tr>
<th>No.</th>
<th>Advantages</th>
<th>Main Idea</th>
<th>Author (Year)</th>
<th>Algorithm</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minimizing energy consumption and computation costs</td>
<td>Dynamic Energy Efficient Resource Allocation in fog environments</td>
<td>Anees Ur Rehman (2020) [40]</td>
<td>DEER</td>
<td>Challenges in adapting to changing network conditions and resource management efficiency</td>
</tr>
<tr>
<td>3</td>
<td>Minimized bandwidth cost and balanced load</td>
<td>Bandwidth cost reduction and load balancing in a cooperative three-layer fog-cloud environment</td>
<td>Mirza M. Maswood (2020) [42]</td>
<td>-</td>
<td>Increased complexity and coordination required among layers</td>
</tr>
<tr>
<td>4</td>
<td>Addressing latency and network bandwidth consumption</td>
<td>Load balancing in fog nodes using particle swarm optimization</td>
<td>D. Baburuao et al. (2021) [43]</td>
<td>Particle Swarm Optimization</td>
<td>Reliance on particle swarm optimization may not suit all scenarios</td>
</tr>
<tr>
<td>6</td>
<td>Decreased energy consumption and resource utilization</td>
<td>Fog-Cluster Based Load-Balancing Technique</td>
<td>Prabhdeep Singh (2022) [45]</td>
<td>-</td>
<td>Dependency on specific architecture of fog clusters</td>
</tr>
<tr>
<td>7</td>
<td>Effective in reducing scheduling latency</td>
<td>RADISH model for dynamic task scheduling in IoT-fog multi-cloud environments</td>
<td>B. Kanbar and K. Faraj (2022) [46]</td>
<td>RADISH</td>
<td>Complexity and computational demands in real-time applications</td>
</tr>
<tr>
<td>8</td>
<td>High resource utilization and reduced response times</td>
<td>Scheduling algorithm using CNN and MPSO for load balancing in fog environments</td>
<td>Fatma M. Talaat (2022) [47]</td>
<td>CNN and MPSO</td>
<td>Complexity and need for accurate data for CNN model training</td>
</tr>
<tr>
<td>9</td>
<td>Improved energy efficiency and resource utilization</td>
<td>Load balancing strategy using fuzzy golden eagle optimization</td>
<td>Simar Preet Singh (2022) [48]</td>
<td>Fuzzy Golden Eagle Optimization</td>
<td>Struggles with dynamic and unpredictable IoT workloads</td>
</tr>
<tr>
<td>10</td>
<td>Efficient smart grid load balancing via fog and cloud computing</td>
<td>VM migration for balancing fog load</td>
<td>Yu. Dongmin (2022) [49]</td>
<td>-</td>
<td>Trade-off between migration costs and load balancing efficiency</td>
</tr>
<tr>
<td>11</td>
<td>Resilience and load balancing in fog networks using a Multi-Criteria Decision Analysis approach</td>
<td>Load balancing in fog networks using ELECTRE</td>
<td>Ebrahim and A. Hafid (2023) [50]</td>
<td>ELECTRE</td>
<td>Difficulties in weighing multiple criteria in variable network conditions</td>
</tr>
<tr>
<td>12</td>
<td>Balancing load and reducing response time</td>
<td>Multi-objective load balancing using a reinforcement learning algorithm in edge-fog-cloud architecture</td>
<td>Fatemeh R. Shahidani (2023) [51]</td>
<td>Reinforcement Learning</td>
<td>Complexity in training and execution</td>
</tr>
</tbody>
</table>
3 Fundamentals of Fog Computing

3.1 Architecture of Fog Computing
Fog computing extends cloud computing by bringing computing resources closer to the edge, near end devices and data sources[54]. It consists of a decentralized architecture spanning across a distributed continuum from cloud to edge devices. Key components include IoT devices, fog nodes (routers, gateways, access points etc.), and cloud data centers. Fog nodes have the capability to process data and execute services, unlike end devices with limited compute ability[55]. But they have lower capacity than the powerful cloud. Multiple tiers of fog nodes are possible based on geographic distribution. This hierarchy improves scalability and reduces service latency.

3.2 Role in IoT Ecosystem
- Fog computing addresses key limitations of cloud regarding Latency, mobility support and geographic distribution. It acts as an intermediate layer for preprocessing IoT data[7].
- Reduces bandwidth requirements to cloud by filtering and temporarily storing data near the network edge. Also enables low latency analytics on streaming data before forwarding to cloud[56].
- Provides location-aware (by using GPS and edge computing for real-time location tracking and rapid data processing) and low latency services for real-time interactions and control of IoT devices. Extends cloud storage/compute services closer to devices[10].

3.3 Key Challenges and Opportunities
Management of heterogeneous devices and dynamic provisioning of compute/network resources across distributed fog nodes makes deployment complex. Lack of common standards and orchestration platforms is an impediment. Vendor specific architectures result in poor interoperability[57]. Security threats are amplified because of wider attack surfaces spanning edge to cloud. Mechanisms for mutual authentication and secure communications are needed[58]. Machine learning approaches need to be integrated for intelligent and self-adaptive resource management to handle variable workloads. Fog supports data accumulation required by some Machine Learning (ML) algorithms too. Scalable software platforms, toolkits for rapid application development and testing, energy-efficiency and commercial viability still pose research opportunities. In summary, fog computing provides a distributed platform bridging end devices to cloud to enable real-time, low latency services and analytics required for future IoT ecosystems. However more standardization is required for greater adoption.

4 Load Balancing in Fog Computing
Load balancing is the process of distributing workload across multiple computing resources to ensure maximum throughput, optimize resource utilization, and avoid overloading of any single resource[20] In fog computing, load balancing helps efficiently utilize the distributed fog nodes that have heterogeneity in terms of compute/storage capacity and network resources. It prevents congestion and bottlenecks especially for latency-sensitive IoT applications. Key benefits include:
- Reduced service latency and improved QoS by directing requests to least loaded fog nodes
- Better scalability as peaked loads can be handled gracefully by transferring excess Load
- Enhanced reliability and fault tolerance based on the redundancies across nodes
- Conservation of energy through optimal resource usage preventing over-provisioning
The following four tables provide an overview of various Load balancing algorithms and techniques that have been proposed for fog computing systems that described in Figure 1, that shows the four fog’s layers. Table 2 focuses on approximate Load balancing algorithms, summarizing the limitations, advantages, main ideas, authors, algorithms, and references for 16 algorithms. Table 3 covers exact Load balancing algorithms, outlining the limitations, advantages, main ideas, authors, algorithms and references for 12 proposed techniques. Table 4 details fundamental Load balancing algorithms for fog systems, highlighting the limitations, advantages, main ideas, authors, algorithms, and references for 12 algorithms. Finally, Table 5 provides an overview of hybrid Load balancing algorithms in fog systems, summarizing the limitations, advantages, main ideas, authors, algorithms, and references for 7 proposed algorithms.

Fig. 1. Fog Computing IoT architecture[59]

<table>
<thead>
<tr>
<th>No.</th>
<th>Algorithm</th>
<th>Author (year)</th>
<th>Main idea</th>
<th>advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>stochastic heuristic (hill climbing algorithm)</td>
<td>M. Zahid et al. (2018) [60].</td>
<td>Hill Climbing Load Balancing Algorithm designed for fog computing environments</td>
<td>1. Local Optimization simplicity</td>
<td>1- Local Optima Trap 2- Dependency on Initial Solution 3- algorithm tend optimized for exploitation rather than exploration</td>
</tr>
<tr>
<td>2</td>
<td>stochastic heuristic</td>
<td>M. B. Kamal et al. (2018) [61]</td>
<td>Load balancing in heuristic min-conflict optimization method</td>
<td>1- efficiency 2- Applicability</td>
<td>1- Sub optimality 2- Sensitivity to Parameters 3- Dependence on Problem Characteristics</td>
</tr>
<tr>
<td>3</td>
<td>stochastic heuristic</td>
<td>X. Xu et al. (2018) [62]</td>
<td>heuristic virtual machine scheduling method designed for load balancing in fog-cloud computing environments</td>
<td>1- Efficiency 2- Adaptability</td>
<td>1- Sub optimality 2- Sensitivity to its setting Parameters</td>
</tr>
<tr>
<td>4</td>
<td>stochastic heuristic</td>
<td>F. Banaie et al. (2020) [63]</td>
<td>Load balancing scheme based on AHP(analytic hierarchy process) method for multiple gateways in a fog computing</td>
<td>The effectiveness of the proposed solution in the fast and reliable acquisition of big data</td>
<td>This algorithm can not address the scenario in which the network has</td>
</tr>
<tr>
<td>No.</td>
<td>Methodology</td>
<td>Reference</td>
<td>Description</td>
<td>Challenges</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Stochastic Meta-heuristics</td>
<td>J. Wan et al. (2018) [64].</td>
<td>Scheduling and load balancing for fog based smart factory</td>
<td>1-Energy Efficiency 2-Improved System Performance 1-Complexity and Overheads 3-Optimality Challenges</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Stochastic Meta-heuristics</td>
<td>J. Yang (2020) [65].</td>
<td>low-latency requirements in the context of healthcare applications, especially those dealing with substantial volumes of medical data</td>
<td>1-Low Latency 2-Improved Data Processing 1-System Complexity 2-Resource Overheads</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Stochastic Meta-heuristics</td>
<td>D. Babbarao et al. (2023) [43].</td>
<td>Resource allocation and load balancing by particle swarm optimization (PSO) in fog environment, optimizing the distribution of computational tasks among fog nodes in a fog computing environment to improve overall system performance</td>
<td>1-Dynamic Adaptation 2-high resource utilization 1-Parameter Sensitivity 2-high complexity 3-Optimization Challenges</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Probabilistic static Machine learning</td>
<td>C. Li et al. (2018) [67].</td>
<td>SSLB (Self-Similarity-Based Load Balancing) designed for large-scale fog computing environments. The focus is on efficiently distributing computational tasks across fog nodes, taking into consideration self-similarity patterns within the workload.</td>
<td>1-1-Efficient Load Balancing 2-Scalability 1-Dependency on Workload Characteristics 2-Algorithm Complexity 3-Performance in Dynamic Environments</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Probabilistic static Game theory</td>
<td>S. F. Abedin et al. (2018) [68].</td>
<td>Fog load balancing method specifically designed for Massive Machine Type Communications (mMTC). The focus is on efficiently distributing the communication load among fog nodes to cater to the requirements of a massive number of machine-type devices in a fog computing environment</td>
<td>1-Efficient Load Balancing 2-Game Theoretic Optimizations 1-High complexity 2-Sensitivity to Parameters 3-Scalability Challenges</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Probabilistic static Fuzzy logic</td>
<td>S. P. Singh et al. (2020) [69].</td>
<td>Design and exploration of load balancers for fog computing using fuzzy logic. The focus is on leveraging fuzzy logic principles to create load-balancing mechanisms tailored for fog computing environments.</td>
<td>1-Adaptability 2-Handling Uncertainty (Fuzzy logic enables robust fog load balancing amid unpredictability) 1-Complexity 2-Interpretability 3-Optimization Challenges</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Probabilistic static Random walk</td>
<td>R. Beraldi et al. (2020) [70].</td>
<td>Fog load balancing algorithm based on random walk (loosely) the algorithm may account for varying and possibly partially correlated states of the fog nodes or the system.</td>
<td>1-Adaptability 2-Scalability 1-Deterministic Challenges 2-Optimization Challenges 3-Resource Overheads</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Algorithm</td>
<td>Author(year)</td>
<td>Main idea</td>
<td>advantage</td>
<td>Limitations</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>--------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>Graph theory Breadth first search</td>
<td>D. Puthal et al. (2018) [21].</td>
<td>Achieving secure and sustainable load balancing for edge data centers in fog computing environments. The emphasis is on optimizing the distribution of computational tasks among edge data centers in a way that ensures both security and sustainability.</td>
<td>high security Sustainability</td>
<td>1-Complexity 2-Resource Overheads 3-Dependence on System Characteristics</td>
</tr>
<tr>
<td>2</td>
<td>Gradient based Newton raphson gauss seidel</td>
<td>E. Barros et al. (2018) [71].</td>
<td>Processing and analyzing real-time data from various sensors and devices within the smart grid, making decisions on load distribution, and responding to changes in the grid's conditions.</td>
<td>1-Real-time Analysis 2-Decentralized Processing 3-low cost</td>
<td>1-Data Security 2-System Complexity 3-Resource Limitations</td>
</tr>
<tr>
<td>3</td>
<td>Gradient based Fixed point algorithm</td>
<td>R. Beraldi and H. Alnuweiri (2018) [72].</td>
<td>Load balancing method based on sequential randomization in fog node, optimizing the distribution of computational tasks among fog nodes in a sequential and randomized manner.</td>
<td>1-Adaptability 2-Scalability</td>
<td>1-Deterministic Challenges 2-Optimization Challenges 3-Resource Overheads</td>
</tr>
<tr>
<td>4</td>
<td>combinational Greedy algorithm</td>
<td>M. Mukherjee et al. (2018) [73].</td>
<td>Load balancing mechanism for F-RAN (Fog Radio Access Networks) optimizing the distribution of computational tasks among fog nodes in a way that considers both the transmission requirements and latency constraints within a radio access network.</td>
<td>Optimized Transmission Reduced Latency</td>
<td>1-Complexity 2-Resource Overheads 3-Sensitivity to Network Conditions</td>
</tr>
<tr>
<td>5</td>
<td>decomposition Linear programming</td>
<td>S. Sthapit et al. (2019) [74].</td>
<td>Explore computational load balancing on the edge in the absence of cloud and fog computing.</td>
<td>1-Edge Resource Utilization 2-Reduced Latency</td>
<td>1-Limited Scalability 2-Resource Heterogeneity 3-Dependency on Edge Infrastructure</td>
</tr>
<tr>
<td>6</td>
<td>Graph theory Dijkstra algorithm</td>
<td>K. Cui et al. (2020) [34].</td>
<td>Load balancing of (USV)an unmanned surface vehicles in a fog system</td>
<td>1- Efficient Task Distribution 2- Enhanced Resource Utilization</td>
<td>1-Environmental Challenges 2-Limited Resources 3-Complexity</td>
</tr>
<tr>
<td>7</td>
<td>Gradient based Gradient algorithm</td>
<td>Q. Fan and N. Ansari (2020) [75]</td>
<td>Workload balancing algorithm in IOT\fog model</td>
<td>1-Efficient Workload Distribution 2-Improved System Performance</td>
<td>1-IoT Device Heterogeneity 2-Communication Overheads, 3-Security and Privacy Concerns</td>
</tr>
<tr>
<td>8</td>
<td>decomposition Weighted sum</td>
<td>M. M. S. Maswood et al.(2020) [42].</td>
<td>Achieve both bandwidth cost reduction and load balancing in a cooperative three-layer fog-cloud computing environment.</td>
<td>1-Bandwidth Cost Reduction 2-Load Balancing</td>
<td>1-Complexity 2-Dependency on Network Conditions 3-Resource Overheads</td>
</tr>
<tr>
<td>No.</td>
<td>Algorithm</td>
<td>Author(year)</td>
<td>Main idea</td>
<td>advantage</td>
<td>Limitations</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>--------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
2-Enhanced Resource Efficiency | 1-Algorithm Complexity  
2-Sensitivity to Parameters  
3-Adaptability to Dynamic Environments |
| 2   | Throttled RR, first fit | N. Ahmad et al. (2019)[77] | Fog-cloud-based platform that utilizes resources efficiently using a load balancing technique. | 1- Resource Utilization  
2- Enhanced Performance | 1-Algorithm Complexity  
2- Dependency on Network Conditions  
3- Scalability Challenges |
| 3   | LBA algorithm | D. A. Chekired et al (2018)[78] | queuing model for Electric Vehicle (EV) energy management and Load balancing method in distributed fog architecture | 1-high scalability  
2-low response time  
3-Efficient Energy Management | 1-Dependency on Fog Infrastructure  
2-Sensitivity to System Dynamics  
3-Dependency on Fog Infrastructure |
| 4   | FoT load balancing algorithm | E. Batista et al. (2018)[79] | address load balancing in Fog of Things (FoT) platform through the utilization of Software-Defined Networking (SDN) | 1-Dynamic Load Balancing  
2-Resource Optimization: by allowing for programmable and centralized control over network configurations | 1-Implementation Challenges  
2-Scalability  
3-Dependency on Network Conditions |
| 5   | Throttled RR | S. Tariq et al (2018)[80] | Load balancing based on priority in fog ‘cloud systems | 1-Task Prioritization  
2-Adaptability | 1-Complexity  
2-Subjectivity of Priority  
3-Resource Overheads: |
| 6   | DEER | A. U. Rehman et al. (2018)[40] | a dynamic energy-efficient resource allocation strategy for load balancing in a fog computing environment. The focus is on optimizing the distribution of computational tasks in a fog environment to enhance energy efficiency dynamically | 1-Energy Efficiency  
2-Load Balancing | 1-Algorithm Complexity  
2-Dependency on Dynamic Conditions |
| 7   | Pricing-based workload distributor algorithm | Z. Sharmin et al (2020)[81] | Achieving sustainable micro-level fog-federated load sharing in the context of the Internet of Vehicles (IoV). The goal is likely to optimize the distribution of computational tasks among fog nodes in a micro-level federated architecture, considering sustainability aspects within the (IoV) ecosystem. | 1-Micro-Level Optimization  
2-Sustainability | 1-Complexity  
2-Dependency on Vehicular Conditions  
3-Real-Time Constraints |
<table>
<thead>
<tr>
<th>No.</th>
<th>Algorithm</th>
<th>Author (year)</th>
<th>Main idea</th>
<th>advantage</th>
<th>limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RR Throttle, ACO</td>
<td>S. A. A. Naqvi et al. (2018) [86]</td>
<td>application of a metaheuristic optimization technique for load balancing in a cloud-fog environment integrated with a smart grid</td>
<td>1-Efficient Load Balancing 2-Integration with Smart Grid</td>
<td>1-high complexity 2-Sensitivity to Parameter Tuning 3-Dependency on Smart Grid Characteristics</td>
</tr>
<tr>
<td>4</td>
<td>RR, Throttle, hybrid genetic algorithm (GA)</td>
<td>M. Zubair et al</td>
<td>using a hybrid genetic algorithm employing bin packing techniques that leverages both</td>
<td>Efficient Load Balancing Adaptable to Dynamic Conditions</td>
<td>1-high complexity</td>
</tr>
</tbody>
</table>

Table 5. Hybrid Load balancing algorithms in the fog system and their properties.
4.1 Impact on QoS

Efficient load balancing has significant impact on improving QoS metrics[92]:

- Throughput: Load distribution prevents congestion and increases number tasks executed per unit time.
- Latency/Response Time: Directing requests to least loaded fog nodes reduces waiting time and speeds up execution.
- Energy Efficiency: Optimized usage of fog nodes saves energy compared to over-provisioning to handle peak loads.
- Reliability: Failover and redistribution of Load in case of node failures improves fault tolerance.

5 Conclusions

The paper extensively explores the impact of Load balancing on QoS in fog computing environments, integral to IoT applications. It identifies various Load balancing algorithms and their efficacy in enhancing system performance. The paper reveals a trajectory from basic optimization models to sophisticated strategies, emphasizing dynamic over static techniques due to their adaptability to fluctuating loads and system changes. Innovations in secure and sustainable Load balancing techniques, such as authenticating Edge Data Centers and allocating tasks to underutilized centers, have been highlighted as significant advancements.

Theoretically, this paper enriches the understanding of fog computing in IoT, particularly in the domain of load balancing. It underscores the necessity of adaptive, secure, and efficient load balancing solutions for the evolving requirements of fog computing environments. Practically, the insights from this paper can guide the development of more efficient IoT systems, particularly in improving QoS through optimized load balancing. The findings have practical relevance in diverse fields like healthcare, smart cities, and vehicular networks, where fog computing plays a pivotal role.

5.1 Limitations of Current Approaches

The paper acknowledges gaps in the literature, particularly in integrating and optimizing Load balancing techniques for the complex scenarios presented by fog computing. It points out the challenges in applying traditional cloud computing load balancing algorithms to the dynamic nature of fog networks. Future research needs to focus on developing algorithms that are more adaptive to the specific requirements of fog computing, addressing issues like scalability, security, and real-time constraints. The paper suggests that further research is required to explore these areas comprehensively, potentially leading to more robust and efficient IoT systems leveraging fog computing.
6 Conclusion

The conclusion of the paper titled "A review Load balancing algorithms in Fog computing", offers a comprehensive summary of its key findings and presents final thoughts on the current state and future prospects of fog computing in the realm of IoT. This paper discusses the algorithms in load balancing that are crucial to improving IoT systems' QoS using fog computing. Many different load balancing algorithms are investigated in depth, and their evolution from traditional load balancers to adaptive, secure, and sustainable load balancers. Such developments are seen as significant in solving the issues here: IoT systems have large amounts of data to deal with. The overall paper illustrates fog computing's current situation in IoT. It acknowledges fog computing as a transformative force in IoT, crucial for handling real-time data processing and reducing Latency. However, it also points out existing challenges in scalability, security, and adaptability to diverse IoT environments. The research advocates for ongoing innovation and exploration in fog computing to fully realize its potential in improving QoS in IoT applications. This call to action underscores the dynamic, ever-evolving nature of fog computing as an essential yet challenging frontier in the IoT landscape, ripe for future research and technological advancements.

Reference


