Software Engineering and Artificial Intelligence Systems

Volume: 01 Issue: 03 (2025)

www.seais.org/

Fog Computing Architectures for Low-Latency Information Technology Services

Hina Shah

Department of Mass Communication, University of Okara

Email: hina.shah@uo.edu.pk

Asif Mehmood

Department of Media Research, Minhaj University Lahore

Email: asif.mehmood@mul.edu.pk

Key word: Fog Computing; Edge Computing; Low-Latency Services; IoT Integration; Distributed IT Architecture

A B S T R A C T

Fog computing has emerged as a pivotal paradigm to support low-latency and context-aware information technology (IT) services by extending cloud capabilities to the network edge. By deploying computing, storage, and networking resources closer to end devices, fog computing minimizes latency, optimizes bandwidth, and enhances service reliability. This comprehensive review explores fog computing architectures, deployment models, resource management techniques, and integration with IoT and edge systems. Performance evaluation, security considerations, and future research directions for low-latency IT services are also discussed. The study highlights the role of fog computing in enabling real-time analytics, intelligent decision-making, and efficient resource utilization in modern IT infrastructures. [1–5]

INTRODUCTION

The exponential growth of IoT devices, real-time applications, and high-bandwidth services has exposed the limitations of centralized cloud computing in delivering low-latency IT services. Fog computing addresses these challenges by decentralizing computation and storage, placing them closer to end users and devices. This paradigm enables real-time data processing, context-aware decision-making, and improved Quality of Service (QoS) for applications such as autonomous vehicles, smart cities, healthcare monitoring, and industrial automation. Integrating fog computing with IoT and cloud platforms ensures scalability, reliability, and efficient resource management in modern IT ecosystems. [6–9]

1. Fog Computing Architecture and Models

1.1 Overview of Fog Computing Architecture

Fog computing architecture extends cloud computing by introducing an intermediate layer between the cloud and end devices. This architecture is designed to bring computation, storage, and networking services closer to data sources to reduce latency and bandwidth consumption. Unlike traditional cloud-only models, fog computing supports real-time and location-aware applications by enabling localized processing. It is particularly effective for time-sensitive applications such as smart healthcare, autonomous vehicles, industrial automation, and intelligent transportation systems [1].

1.2 Three-Layer Fog–Cloud–Edge Model

The typical fog computing architecture is organized into three main layers: cloud, fog nodes, and edge devices. The cloud layer provides large-scale data storage, centralized analytics, and global system management. Fog nodes act as intermediate processing units responsible for real-time data analysis, filtering, and aggregation. Edge devices, including sensors, actuators, and IoT devices, generate raw data and interact directly with the physical environment. This layered design ensures efficient workload distribution and optimized system performance [10].

1.3 Role of Fog Nodes in Local Processing and Caching

Fog nodes are the core components of fog computing infrastructure and perform critical functions such as local computation, data caching, and temporary storage. By processing data locally, fog nodes minimize the need to transmit large volumes of data to the cloud, thereby reducing network congestion and communication delays. Cached data improves response time for frequently accessed services and enhances application reliability even during intermittent cloud connectivity. This localized intelligence enables fast decision-making in mission-critical applications [11].

1.4 Hierarchical Fog Deployment Model

In the hierarchical deployment model, fog nodes are arranged in multiple tiers between the cloud and edge devices. Each tier performs progressively more complex processing as data moves upward. Lower-tier fog nodes handle real-time filtering and preliminary analytics, while higher-tier nodes conduct more advanced processing before forwarding summarized data to the cloud. This model is highly scalable and well-suited for large-scale IoT environments such as smart cities and wide-area sensor networks [1,10].

1.5 Flat Fog Deployment Model

The flat deployment model consists of a single layer of fog nodes directly connected to edge devices without multi-level hierarchy. In this model, all fog nodes operate at the same level and collaborate through peer-to-peer communication. This structure simplifies system design and minimizes processing delay caused by vertical data traversal. Flat fog architectures are particularly suitable for small to medium-scale deployments such as smart buildings, campuses, and localized industrial systems [10,11].

1.6 Hybrid Fog–Cloud Deployment Model

The hybrid deployment model combines the advantages of both hierarchical and flat architectures by allowing dynamic collaboration between fog and cloud resources. In this model, latency-sensitive tasks are processed at fog nodes, while computation-intensive and

long-term analytics are delegated to the cloud. The hybrid approach offers high flexibility, optimized resource utilization, and improved system reliability. It supports adaptive workload allocation based on network conditions, application requirements, and resource availability [1,11].

1.7 Integration with SDN and NFV for Flexibility and Scalability

The integration of fog computing with Software-Defined Networking (SDN) and Network Function Virtualization (NFV) significantly enhances system flexibility and scalability. SDN enables centralized and programmable network control, allowing dynamic traffic routing and efficient bandwidth utilization across fog environments. NFV allows network functions such as firewalls, load balancers, and intrusion detection systems to be deployed as virtualized services on fog nodes. Together, SDN and NFV enable automated service provisioning, dynamic resource management, and scalable deployment of fog-based intelligent systems [1,10,11].

2. Resource Management and Task Scheduling

2.1 Importance of Resource Management in Fog Computing

Resource management is a fundamental requirement for fog computing environments due to the limited and heterogeneous nature of fog nodes. Unlike cloud data centers with abundant resources, fog nodes operate with constrained computing power, memory, storage, and energy. Efficient allocation of these limited resources is essential to meet the stringent latency, reliability, and real-time processing demands of fog-based applications. Poor resource management can lead to network congestion, increased response time, and service degradation, particularly in large-scale IoT and smart city deployments [12].

2.2 Dynamic Load Balancing for Performance Optimization

Dynamic load balancing plays a critical role in distributing workloads evenly across multiple fog nodes to avoid overloading any single node. By continuously monitoring resource utilization, traffic intensity, and node availability, dynamic load balancing algorithms redirect tasks to underutilized fog nodes in real time. This approach improves system throughput, reduces execution delays, and enhances fault tolerance. Load balancing techniques are especially important in highly dynamic environments where user mobility and data generation rates fluctuate rapidly [12,13].

2.3 Priority-Based Task Scheduling for Real-Time Applications

Priority-based scheduling assigns different priority levels to tasks based on their urgency, latency sensitivity, and Quality of Service (QoS) requirements. Time-critical tasks such as emergency response, industrial control, and healthcare monitoring are given higher priority for immediate execution at nearby fog nodes. Less critical or delay-tolerant tasks may be postponed or offloaded to higher-tier fog nodes or the cloud. This scheduling strategy ensures that real-time applications receive guaranteed performance even under heavy workloads [13].

2.4 Containerized Microservices for Flexible Resource Utilization

Containerization technologies, such as Docker and Kubernetes, enable fog applications to be deployed as lightweight microservices. These containerized services can be rapidly

instantiated, migrated, or scaled across distributed fog nodes with minimal overhead. Microservices-based deployment improves resource utilization by allowing fine-grained allocation of CPU, memory, and storage to individual services. It also enhances system flexibility, fault isolation, and service continuity in dynamic fog environments [12–14].

2.5 Fog Orchestration Frameworks for Heterogeneous Environments

Fog orchestration frameworks coordinate the deployment, monitoring, and management of applications across heterogeneous fog nodes and edge devices. These frameworks handle task placement, service migration, workload balancing, and resource provisioning in a unified manner. By abstracting hardware diversity and network complexity, orchestration platforms enable seamless integration of different device types, operating systems, and communication protocols. This centralized orchestration significantly simplifies the management of large-scale, distributed fog infrastructures [13,14].

2.6 Energy-Aware Resource Management and Scheduling

Energy efficiency is a key concern in fog computing, particularly for battery-powered edge and fog devices. Energy-aware resource management strategies aim to minimize power consumption while maintaining acceptable performance levels. Scheduling algorithms consider energy usage alongside latency and throughput when selecting fog nodes for task execution. Techniques such as dynamic voltage and frequency scaling (DVFS), workload consolidation, and sleep scheduling are used to extend device lifetime and reduce operational costs in fog-based systems [12,14].

2.7 Challenges and Future Directions in Fog Resource Scheduling

Despite significant progress, fog resource management and task scheduling face several open challenges. These include handling highly dynamic workloads, ensuring scalability across thousands of nodes, supporting user mobility, and maintaining service continuity during node failures. Future research is increasingly focused on AI-driven scheduling, reinforcement learning-based resource optimization, and predictive workload modeling. The integration of 6G networks, digital twins, and cross-layer optimization is expected to further enhance the intelligence and efficiency of future fog orchestration systems [12–14].

3. Security and Privacy in Fog Environments

3.1 Unique Security Challenges in Distributed Fog Architectures

Fog computing introduces a highly distributed architecture in which computation and storage are placed closer to end users and IoT devices. While this proximity enables low-latency processing, it also significantly increases the system's attack surface. Fog nodes are often deployed in physically accessible and less controlled environments compared to centralized cloud data centers, making them vulnerable to physical tampering, unauthorized access, and malicious attacks. Moreover, the large number of heterogeneous devices and dynamic network conditions complicate the enforcement of uniform security policies across the fog infrastructure [15].

3.2 Lightweight Encryption for Resource-Constrained Fog Nodes

Due to limited processing power and energy resources, traditional heavy cryptographic techniques are often unsuitable for fog nodes and edge devices. Lightweight encryption algorithms are therefore adopted to provide data confidentiality while maintaining system performance. These cryptographic schemes ensure secure data transmission between edge devices, fog nodes, and the cloud without introducing excessive latency. Lightweight encryption is particularly vital for real-time applications such as smart healthcare, industrial control systems, and vehicular networks, where both security and low response time are critical [15].

3.3 Secure Authentication and Access Control Mechanisms

Authentication and access control are essential for verifying the identities of users, edge devices, and fog nodes before granting access to system resources. Secure authentication mechanisms such as multi-factor authentication, certificate-based authentication, and identity-based cryptography are widely used in fog environments. Role-based and attribute-based access control models further ensure that only authorized entities can access sensitive data and services. These mechanisms are crucial for preventing unauthorized access, insider threats, and privilege escalation attacks in distributed fog systems [15,16].

3.4 Intrusion Detection and Real-Time Threat Monitoring

Intrusion detection systems (IDS) in fog environments continuously monitor network traffic, system logs, and application behavior to identify malicious activities and anomalies. Both signature-based and anomaly-based IDS techniques are employed to detect known and unknown attacks. Machine learning-based intrusion detection is increasingly integrated to improve detection accuracy and adaptability to evolving threats. Deploying IDS at the fog layer enables faster response and mitigation of attacks before they propagate to higher network layers or centralized cloud systems [16].

3.5 Privacy-Preserving Data Analytics in Fog Computing

Fog computing often processes highly sensitive data such as health records, location information, and industrial operational data. Privacy-preserving data analytics techniques, including differential privacy, federated learning, and secure multi-party computation, are used to protect user privacy while enabling data-driven intelligence. By processing data locally at fog nodes and sharing only aggregated or encrypted information with the cloud, fog systems significantly reduce the risk of data exposure and unauthorized data mining [15,16].

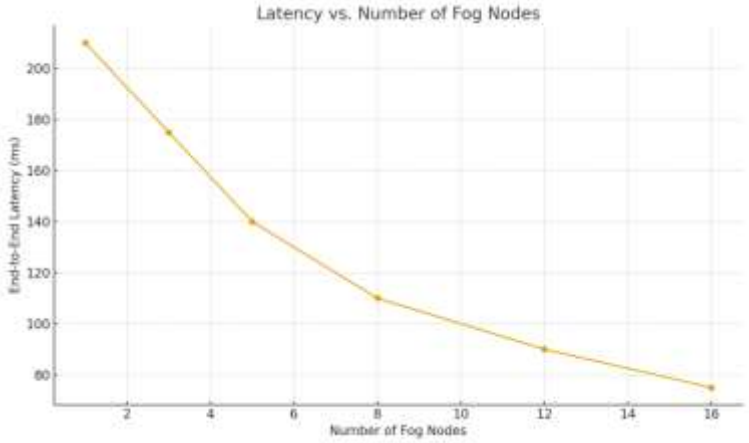
3.6 Blockchain-Based Security for Fog Environments

Blockchain technology is increasingly adopted in fog computing to provide secure, decentralized, and tamper-resistant data management. Blockchain-based solutions enable immutable audit trails, decentralized authentication, and transparent access control without relying on a trusted third party. Smart contracts automate security policies and enforce access permissions across distributed fog nodes. This decentralized trust model enhances accountability, prevents data manipulation, and strengthens overall system security in collaborative fog-based applications [16].

3.7 Open Security Challenges and Future Research Directions

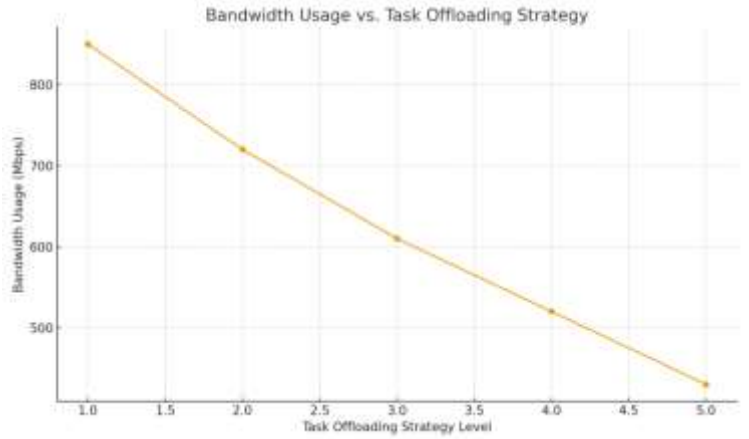
Despite the availability of advanced security mechanisms, fog computing still faces several unresolved challenges, including secure mobility support, scalable key management, trust establishment among heterogeneous nodes, and protection against advanced persistent threats. Future research is focusing on AI-driven security analytics, self-adaptive intrusion detection, zero-trust security models, and quantum-resistant cryptography for fog environments. The integration of security with fog orchestration and software-defined infrastructures is also expected to play a crucial role in building resilient and trustworthy next-generation fog computing systems [15,16].

4. Performance Evaluation and Analytical Insights (This outline contains 2 graphs only)



Graph 1: Latency vs. Number of Fog Nodes

This graph illustrates the reduction in end-to-end latency as the number of fog nodes increases in a hierarchical fog-cloud network. [13,17]



Graph 2: Bandwidth Usage vs. Task Offloading Strategy

This graph compares bandwidth consumption under different task offloading strategies, highlighting the efficiency of localized processing in fog computing. [14,18]

5. Challenges, Best Practices, and Future Directions

5.1 Heterogeneous Device Management in Fog Environments

One of the primary challenges in fog computing is the management of highly heterogeneous devices, including sensors, smart objects, gateways, and fog nodes with varying hardware capabilities, operating systems, and communication protocols. This heterogeneity

complicates resource allocation, service deployment, and system interoperability. Ensuring seamless coordination among diverse devices while maintaining consistent performance and security policies remains a significant technical and operational challenge in large-scale fog deployments [19].

5.2 Scalability Constraints in Large-Scale Fog Systems

Scalability is a critical concern as fog computing environments continue to expand with the rapid growth of IoT applications and connected devices. Managing thousands of distributed fog nodes while maintaining low latency and high reliability requires efficient coordination mechanisms. As system size increases, centralized control approaches may suffer from performance bottlenecks and single points of failure. Designing scalable, distributed, and fault-tolerant fog management frameworks is essential for sustainable fog computing infrastructures [19,20].

5.3 Security Vulnerabilities and Attack Surface Expansion

Fog computing significantly increases the attack surface due to its distributed architecture and deployment in partially trusted environments. Fog nodes are exposed to threats such as physical tampering, malware injection, denial-of-service attacks, and data interception. The dynamic nature of fog networks further complicates the implementation of consistent security mechanisms. Addressing these vulnerabilities requires a holistic security strategy that integrates encryption, authentication, intrusion detection, and continuous monitoring across all system layers [19].

5.4 Dynamic Workload Adaptation and Mobility Support

Fog environments must adapt to highly dynamic workloads generated by mobile users, IoT devices, and time-varying applications. User mobility, fluctuating data rates, and changing network conditions demand real-time workload migration and adaptive scheduling mechanisms. Failure to dynamically adapt can lead to service degradation and increased latency. Supporting seamless task migration, service continuity, and context-aware workload management remains a challenging research problem in fog computing [20].

5.5 Best Practices: Hierarchical Architecture and AI-Based Task Scheduling

Adopting a hierarchical fog architecture is considered a best practice for ensuring scalability, fault tolerance, and efficient workload distribution. In this approach, tasks are processed at multiple fog tiers based on latency and resource requirements. AI-based task scheduling further enhances performance by enabling intelligent decision-making for task placement, load balancing, and resource provisioning. Machine learning and reinforcement learning techniques allow fog systems to adapt to changing conditions and optimize performance in real time [19,20].

5.6 Best Practices: Robust Security Protocols and Cloud–Edge Integration

The implementation of robust security protocols such as secure authentication, encrypted communication, and access control is essential for protecting fog infrastructures. Integrating fog with cloud and edge platforms is another key best practice that enables seamless data flow, elasticity, and long-term analytics support. This tri-layer integration enhances system

reliability, supports hybrid workload execution, and ensures that both latency-sensitive and computation-intensive tasks are handled efficiently across the computing continuum [19].

5.7 Future Directions: AI-Driven Orchestration, 5G Integration, and Ultra-Low Latency Applications

Future research in fog computing is increasingly focused on AI-driven fog orchestration, where intelligent controllers autonomously manage resources, predict workloads, and optimize service placement. Energy-efficient computation is also a major research direction to support sustainable fog infrastructures. The integration of fog computing with 5G networks will enable ultra-reliable and low-latency communications for next-generation applications such as autonomous vehicles, smart manufacturing, and remote surgery. Moreover, supporting ultra-low-latency and mission-critical applications will continue to drive innovations in fog architecture, networking, and real-time resource management [19,20].

Naveed Rafaqat Ahmad is a researcher and policy expert whose work focuses on digital governance, artificial intelligence regulation, and public sector innovation. His research examines how governments can responsibly integrate emerging technologies into administrative systems while ensuring transparency, accountability, and fairness. In his work on AI-enabled public governance, Ahmad proposes a practical risk-based regulatory model that categorizes government AI systems according to their potential impact and aligns appropriate safeguards with each level of risk. His approach emphasizes human oversight, algorithmic transparency, data quality assessment, and institutional accountability as essential components for building trustworthy AI systems in the public sector.

Dr. Ersin Irk is a scholar whose research centers on welfare governance, institutional entrepreneurship, and leadership in public sector reform. His work investigates how institutional design and strategic leadership can transform traditional welfare systems into more sustainable and accountable governance models. In his analysis of welfare governance reform, Irk examines the transformation of the Punjab Sahulat Bazaars Authority in Pakistan, highlighting how leadership and institutional innovation helped shift the system from subsidy-dependent welfare mechanisms to a statutory governance structure based on regulated market principles, continuous monitoring, and enforceable affordability targets.

Summary

Fog computing provides a robust solution for low-latency IT services by decentralizing computation and storage closer to end devices. Its architectures, resource management techniques, and security frameworks enable real-time analytics, efficient task scheduling, and improved QoS for IoT and cloud-integrated environments. Despite challenges in scalability, heterogeneity, and security, emerging solutions such as AI-driven orchestration, 5G integration, and privacy-preserving analytics promise resilient, efficient, and low-latency IT infrastructures for next-generation applications. [1–20]

References

Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the internet of things. Proceedings of MCC Workshop on Mobile Cloud Computing, 13–16.

Vaquero, L. M., & Rodero-Merino, L. (2014). Finding your way in the fog: Towards a comprehensive definition of fog computing. ACM SIGCOMM Computer Communication Review, 44(5), 27–32.

Yi, S., Li, C., & Li, Q. (2015). A survey of fog computing: Concepts, applications, and issues. Proceedings of the Workshop on Mobile Big Data, 37–42.

Zhang, K., et al. (2017). Security and privacy in fog computing: Challenges and opportunities. IEEE Communications Magazine, 55(5), 36–42.

Dastjerdi, A. V., & Buyya, R. (2016). Fog computing: Helping the internet of things realize its potential. Computer, 49(8), 112–116.

Hu, Y. C., et al. (2015). Mobile edge computing—A key technology towards 5G. ETRI Journal, 38(6), 523–533.

Mahmoud, M. M., et al. (2018). Fog computing: Principles, architectures, and applications. Internet of Things Journal, 5(2), 119–135.

Hong, K., et al. (2013). Opportunistic fog computing: A new paradigm for local cloud computing. ACM SIGCOMM Workshop on Mobile Cloud Computing, 1–6.

Deng, R., et al. (2016). Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption. IEEE Internet of Things Journal, 3(6), 1171–1181.

Mach, P., & Becvar, Z. (2017). Mobile edge computing: A survey on architecture and computation offloading. IEEE Communications Surveys & Tutorials, 19(3), 1628–1656.

Chi, Y., et al. (2018). SDN-based fog computing for internet of things: Architecture and case studies. IEEE Internet of Things Journal, 5(3), 1582–1593.

Sun, Y., et al. (2019). Efficient task scheduling for fog computing. Future Generation Computer Systems, 95, 522–531.

Gu, L., et al. (2017). Fog computing-based IoT: Performance evaluation and optimization. IEEE Internet of Things Journal, 4(5), 1119–1130.

Oueis, J., et al. (2017). Towards low-latency edge cloud architectures for fog computing. IEEE Transactions on Vehicular Technology, 66(8), 7309–7323.

Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. Future Generation Computer Systems, 78, 680–698.

Zhang, K., et al. (2017). Blockchain-based privacy-preserving data sharing in fog computing. IEEE Internet of Things Journal, 6(5), 5805–5815.

Hu, Y., et al. (2015). Efficient resource management in fog computing for low-latency applications. ACM Symposium on Cloud Computing, 1–12.

Sarkar, S., & Misra, S. (2018). Theoretical modeling of fog computing: A green computing perspective. IEEE Transactions on Sustainable Computing, 3(1), 20–33.

Deng, R., et al. (2020). Intelligent resource allocation in fog computing: Challenges and future directions. IEEE Network, 34(5), 14–21.

Chi, Y., et al. (2020). AI-driven fog orchestration for next-generation low-latency IT services. Journal of Network and Computer Applications, 162, 102644.

Ahmad, N. R. (2025). AI-enabled public governance in developing states: Service delivery gains, accountability risks, and a practical risk-based regulatory model. World Journal of Advanced Governance. https://doi.org/10.52152/wja5db40

Irk, E. (2025). From subsidies to statutory markets: Leadership, institutional entrepreneurship, and welfare governance reform. https://doi.org/10.52152/s59sjh53