iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81077

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Edge Computing vs Cloud Computing: A Comparative Review of Performance, Security, and Scalability

Bandaru Prasanna[1], Saladi Gayathri Devi, Yalla Lavanya[3]
Department of Computer Applications, Aditya University, Surampalem, India,

Abstract: The explosion of IoT devices, real-time apps, and bandwidth-hungry services has forced a hard look at how we've traditionally relied on cloud computing. The cloud is great for centralized, scalable infrastructure that doesn't break the bank—but its latency and bandwidth limits are becoming real problems for applications that need fast responses or span lots of locations.

That's where edge computing comes in, moving processing power closer to where data is actually generated and used. This paper compares edge and cloud computing across three core areas: performance, security, and scalability. Drawing on existing research and architectural analysis, it examines how each approach handles the demands of things like autonomous vehicles, smart healthcare, industrial automation, and augmented reality. The takeaway: edge computing wins on latency and bandwidth efficiency, while cloud computing still dominates when you need sheer computational muscle and global reach. Hybrid architectures— blending edge and cloud—look like the most promising path forward, combining the best of both worlds. The goal here is to give researchers, system architects, and practitioners a clearer picture when designing the next generation of distributed systems.

Keywords: Edge Computing; Cloud Computing; Latency; Scalability; Security; IoT; Hybrid Architecture

I. INTRODUCTION

A. Background

The widespread use of networked devices worldwide has completely reshaped the data creation and consumption environment. As per the recent industry analyses, the world is expected to have more than 30 billion IoT devices by 2030 [1]. The legacy cloud computing solutions that are based on centralized data centers hundreds or thousands of miles away did not have the capacity to support the sheer amount, speed and diversity of data created at the edge of the network. The latency requirements of applications in which a near-instantaneous response is required, such as autonomous driving systems, real-time medical monitoring systems, and industrial control systems, demonstrate the weaknesses of cloud-only systems, where even a few hundred milliseconds of latency can make a service useless or even unsafe.

Cloud computing appeared at the beginning of the 21 st century as a revolutionary technology that allowed organizations to outsource their processing and storage capabilities to huge, remotely operated information centres. This model democratized the use of strong computing resources, enabling both large and small businesses to expand quickly without having to spend a lot of capital [2]. Nonetheless, with changing application needs, there has been a growing concern that the architectural assumptions of cloud computing are being questioned.

B. Problem Statement

In spite of the strongly proven benefits of the cloud computing, a number of challenges remain. One of the underlying issues of time-sensitive applications is network latency. Sending raw sensor data between thousands of geographically distributed devices to a centralized cloud will introduce propagation delays, require large amounts of bandwidth and introduce potential bottlenecks that can slow application responsiveness.

Moreover, the uninterrupted transfer of sensitive information, such as health indicators, financial activities, and surveillance images, to distant servers can bring significant privacy and regulatory issues, especially in the framework of regulations like the General Data Protection Regulation (GDPR) [3].

Edge computing has been suggested as a remedy to such complications by shifting the computational activities in the centralized data centers to the devices at or close to the source of the data. Nevertheless, edge computing does have its flaws such as limited processing power, storage, complicated device operations, and a disjointed security ecosystem.

Knowing when to use edge solutions, when to use the cloud, and how to create successful hybrid architectures is thus a key challenge to the computing community.

C. Objective

The purpose of this paper is to offer a detailed and fair comparative analysis of the edge computing and cloud computing. In particular, it aims to: (i) describe architectural differences between the two paradigms; (ii) compare their respective performance characteristics in terms of latency, bandwidth, and processing capacity; (iii) compare their security models and vulnerabilities; (iv) consider their scalability, both technically and economically; and (v) discuss the areas of their relevance and future research directions. The analysis is based on a broad range of recent scholarly sources to make sure that conclusions are up-to-date with the field.

II. LITERATURE REVIEW

The academic literature has been quite active in conducting comparative analysis of edge and cloud computing. Shi et al. [4] presented the notion of edge computing as a specific paradigm and clearly stated its benefits in terms of latency minimization of IoT applications. Their initial research defined the vocabulary and conceptualization on which other researchers have constructed their work. The authors claimed that taking computation to the network edge would help lighten the load on the backbone infrastructure and also provide real-time analytics at the point of data creation.

The seminal definition of cloud computing created by Mell and Grance [2] to the National Institute of Standards and Technology (NIST) defined cloud computing as a framework that facilitated ubiquitous and convenient access to a pool of configurable computing resources on a pay-as-you-use basis. This definition has been the canonical point of reference on cloud computing studies. Future research, including that of Armbrust et al. [5], investigated the financial nature of cloud computing and proved that the pay-as-you-go nature provides significant cost benefits over having dedicated on-premises infrastructure, especially when the workload of an organization is not predictable on a daily basis.

Security implications of both architectures have been extensively discussed. Roman et al. [6] observed security as one of the leading issues in edge computing environments, and observed that the distributed nature of edge nodes increases the size of the attack surface over that of a centralized cloud. They suggested the following edgespecific taxonomy of security threats: node compromise, man-in-the-middle attacks, and the insertion of rogue devices. Conversely, a survey conducted by Subashini and Kavitha [7] on the security issue of cloud computing identified the risks of multi-tenancy, insider attacks, and data breaches in giant data centers.

A number of researchers have studied the performance tradeoffs of the two models in particular application environments. Satyanarayan et al. [3] investigated the appropriateness of edge and cloud architectures to vehicular networks and found that edge nodes placed at roadways could respond to safety critical applications in an order of magnitude less time than a cloud-only solution. On the same note, Hassan et al. [8] examined how to deploy machine learning inference to the edge, discovering that edge-based inference could deliver 60-80% latency performance improvements compared to cloud-based inference on computer vision tasks in smart surveillance systems.

More recent research has been on hybrid architectures which attempt to integrate the low latency of edge computing with the scalability and computational depth of the cloud. The idea of a middle ground between edge and cloud expressed by Bonomi et al. [9] as a way of distributing the workloads intelligently over the network hierarchy was termed as fog computing. Their framework has inspired further suggestions of task offloading algorithms, dynamic resource management and federated learning structures. The new wave of ideas in the literature is that no one paradigm proves to be superior and in fact the best architecture is the one that matches the needs of the application field in question [10].

III. SYSTEM OVERVIEW AND ARCHITECTURE

It is important to find out the underlying architectural distinctions between edge computing and cloud computing so that their capabilities and limitations can be judged. Figure 1 illustrates schematically the two architectures and how they interact in a contemporary computing ecosystem.

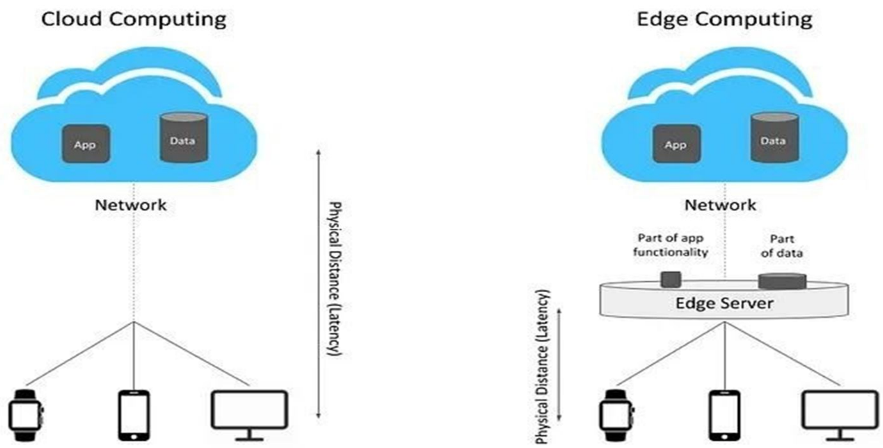


Fig. 1. Edge vs Cloud Computing Architecture

A. Cloud Computing Architecture

The model on which cloud computing is based is centralization. Providers like Amazon Web Services, Microsoft Azure, and Google Cloud platform run large data centers, which contain thousands of interconnected servers that collaboratively create a common pool of computation and storage resources. This pool is accessed by end users and applications via the public internet where virtual machines, managed databases, object storage and an assortment of an extensive range of platform services are on demand. The design is that of huge economies of scale, where sharing the resources with millions of customers, cloud services are able to provide the computational resources at a fraction of the price of the equivalent dedicated hardware.

Cloud platforms are normally architecturally using a multi-tier model. The interconnection of a global network of data centers is through high bandwidth private backbones whereby the regional availability zone offers redundancy and disaster recovery services. The technologies of virtualization and containerization enable the dynamic provisioning, migration, and scaling of workloads based on demand. Kubernetes and other orchestration systems are an automated deployment and management framework of containerized applications, allowing highly scaleable application deployment. The outcome is a platform which can support huge, geographically distributed workloads at high availability and fault tolerance.

B. Edge Computing Architecture

Edge computing decentralizes computing to nodes at, or close to, the physical edge of the network-proximate to the data sources and end users they support. Practically, edge nodes can be in the form of micro data centers placed in cellular base stations, industrial gateways installed on factory floors, or ruggedized computing devices that are placed in cars. Its peculiar feature is physical and logical closeness to the point of data creation that essentially shortens the time spent to transfer data to a computing resource and to transfer results to the application.

An edge deployment architecture is heterogeneous in nature. Computational capability of edge nodes can range from simple microcontrollers with resources to support simple sensor preprocessing to high-performance multicore servers, with the capability to do complex analytics and machine learning inference. These nodes are normally orchestrated through an edge management layer which manages the discovery of devices, task scheduling, monitoring and enforcement of security policy. Most edge deployments are in a consistent connection to a cloud backend managing tasks that are too large to run on the edge, as well as global data aggregation, and long-term storage and analytics.

C. Key Architectural Differences

The most significant architectural variations of the two paradigms are associated with where the computations are done, the control model and the patterns of data flows. In cloud computing, all significant computation is done in remote data centers; generated data at the edge needs to go over the public internet to be processed. In edge computing, a significant portion of computing is done on-site, with only particular data or calculated outputs sent to the cloud. This data flow reversal has far-reaching consequences in the aspects of latency, bandwidth usage, and privacy.

In terms of control, cloud computing is biased towards a centralized control model, in which one orchestration layer controls all the resources in the network of a provider.

By contrast, edge computing will need distributed management; an edge node or node cluster will need to have the capability to operate autonomously, as the network connectivity may not be constant or reliable, and therefore connectedness to a central controller may be unavailable. This independence enhances resistance to network outages at the cost of coordinating activities on a system-wide basis, updating software, and implementing security policy.

IV. COMPARATIVE ANALYSIS

This section provides a systematic comparison of edge computing and cloud computing based on the main performance dimensions that are of the utmost importance to the modern application deployments. The main differences are outlined in Table 1, and the discussion of each parameter is provided.

Table 1. Comparison of Edge Computing and Cloud Computing

Parameter	Edge Computing	Cloud Computing
Latency	Very low (1–10 ms); processing near the source reduces round-trip delays	Moderate to high (50–150 ms); dependent on distance to centralized servers
Processing Power	Limited; relies on lightweight edge nodes and IoT gateways	Virtually unlimited; vast server farms with scalable compute resources
Security	Distributed attack surface; local data handling reduces cloud exposure	Centralized control with mature security protocols but a high-value target
Scalability	Horizontal scaling across many edge nodes; complex to orchestrate	Highly elastic; resources provisioned and deprovisioned on demand
Bandwidth Usage	Low; raw data is filtered locally before selective cloud upload	High; large volumes of raw data transmitted continuously to the cloud
Cost Model	Higher upfront hardware investment; lower operational bandwidth cost	Pay-as-you-go subscription model; can escalate with large data volumes

A. Latency and Response Time

Latency is perhaps the most frequently cited differentiator between edge and cloud computing. In cloud deployments, end-to-end latency is determined primarily by the physical distance between the user or device and the cloud data center, as well as the number of network hops traversed. Typical round-trip times between a mobile device and a public cloud endpoint range from 50 to 150 milliseconds under normal conditions, and may increase substantially during periods of network congestion or when inter-continental routing is required. For applications such as interactive augmented reality, haptic feedback systems, or autonomous vehicle control, these latencies are operationally unacceptable.

Edge computing dramatically reduces this latency by placing computation within one or two network hops of the end device. Measurements reported in the literature indicate that edge-enabled deployments can achieve response times in the range of 1 to 10 milliseconds for local processing tasks [4]. This order-of-magnitude improvement is transformative for latency-sensitive applications and constitutes the primary motivation for edge deployments in industrial, automotive, and healthcare contexts.

B. Processing and Storage Capacity

Cloud computing enjoys an overwhelming advantage in raw processing and storage capacity. The economies of scale achievable in large data centers, combined with the ability to aggregate resources across geographically distributed facilities, mean that cloud platforms can offer virtually unlimited computation on demand. GPU clusters, high-memory instances, and specialized accelerators such as tensor processing units are available as metered services, enabling computationally intensive workloads—including deep learning training, large-scale genomic analysis, and climate simulation—that would be impossible to execute at the edge.

Edge nodes, by contrast, operate under significant resource constraints. The hardware deployed at the edge must balance computational capability against size, weight, power consumption, and cost considerations that are far less pressing in a data center environment. While purpose-built edge inference accelerators have improved significantly in recent years, edge deployments remain better suited to inference, data filtering, and lightweight analytics than to the training of large-scale machine learning models or the processing of petabyte-scale datasets.

C. Security Considerations

Security represents a domain where neither paradigm holds an unambiguous advantage. Cloud computing benefits from decades of investment in centralized security infrastructure, including perimeter firewalls, intrusion detection systems, multi-factor authentication, and sophisticated access control frameworks. The concentration of resources and expertise within a small number of well-funded providers means that the security posture of major cloud platforms is often superior to what individual organizations could achieve independently. However, this centralization also makes cloud data centers high-value targets for sophisticated adversaries, and a successful breach can expose the data of millions of customers simultaneously.

Edge computing distributes data across many geographically dispersed nodes, which limits the blast radius of any individual breach and reduces the exposure of sensitive data to transit over the public internet. However, the same distribution that limits the impact of a single compromise also multiplies the number of potential attack surfaces. Edge nodes are frequently deployed in physically insecure environments, where they may be susceptible to tampering, theft, or side-channel attacks. The heterogeneity of edge hardware and software stacks complicates the application of uniform security policies and makes patch management significantly more challenging than in a homogeneous cloud environment [6].

D. Scalability

Cloud computing is architecturally designed for scalability. The elasticity of cloud resources—the ability to provision and release computational capacity in near real time in response to demand—is one of its defining characteristics. Auto-scaling groups, serverless computing frameworks, and globally distributed content delivery networks allow cloud-based applications to accommodate demand spikes of several orders of magnitude without manual intervention. This scalability is available as a service, relieving application developers of the need to anticipate and provision for peak loads.

Scaling an edge deployment is a more complex undertaking. Adding computational capacity at the edge requires the physical deployment of additional hardware in potentially remote or hostile environments, coordination of device management software, and integration with existing network infrastructure. Horizontal scaling across large numbers of edge nodes introduces challenges in maintaining consistency, managing distributed state, and routing requests intelligently. These challenges are not insurmountable, but they represent a substantially higher operational burden than equivalent cloud scaling operations

.

V. APPLICATIONS AND USE CASES

A. Real-Time Systems

Perhaps the most interesting applications for edge computing involve real-time response. For example, in a selfdriving vehicle, the car needs to process information from tens of sensors, including cameras, LiDAR, radar, and ultrasound, and make safety-critical decisions and take action in milliseconds. The latency of a round trip to a data center is too great to support these demands, especially in degraded cellular conditions. The edge computing paradigm allows for processing on the vehicle or at the side of the road to minimise latency to the point where safe autonomous driving is possible [3].

Industrial automation and robotics systems with closed-loop control must also have sub-millisecond response time. For example, the response time required of a robotic arm performing complex assembly, a CNC tool making a tight tolerance cut or a conveyor system responding to an error in sensor data is in the hundreds of microseconds, not milliseconds. Cloud-connected edge controllers on the shop floor supply the local processing power needed for local control, as well as connections to the cloud for data capture, predictive maintenance, and fleet management.

B. Smart Infrastructure and healthcare

In medical applications, edge computing can be used for real-time patient monitoring via wearable sensors, with local alarms and anomaly detection that can be performed offline. This is especially useful in environments like operating rooms, ICUs or remote clinics where connectivity is not always reliable. This also minimises the amount of sensitive data that must be sent to the cloud, addressing privacy concerns and enabling regulatory compliance with standards like the Health Insurance Portability and Accountability Act (HIPAA) and equivalent standards in other countries.

The data generated by smart city infrastructure such as traffic control and environmental monitoring sensors and public safety surveillance cameras is vast and would overwhelm cloud resources if it were to be sent in full. Edge processing units installed at intersections or on street lighting can do simple filtering, object recognition and event detection locally and only report relevant events or statistical summaries to the cloud. This not only reduces the cost of transmitting data over the network, but also enables lower latency and continuous operation even when not connected to the cloud.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

C. Cloud-Native Applications

When it comes to applications with dynamic or unpredictable workloads, worldwide user populations, and heavy processing demands, cloud computing remains the ideal platform. The elasticity, global reach and managed services provided by public cloud providers are a boon to Software-as-a-Service (SaaS) applications, commerce platforms, media streaming, and collaboration and productivity applications. The capacity to service users on multiple continents from a limited number of geographically distributed data centers, along with advanced content delivery networks, can offer users an experience that is difficult to match with edge computing.

VI. CHALLENGES

A. Challenges in Edge Computing

There are a number of challenges that limit the use of edge computing. Limited resources are perhaps the most obvious: edge nodes are often constrained by a trade-off between computational throughput and other constraints such as energy consumption, physical footprint, thermal properties, and cost; the latter often results in in trade-off for capabilities compared to data center nodes. Distributed management of an edge system is considerably more challenging than a homogeneous cloud system; hardware diversity, environmental variability, and physical inaccessibility make software upgrades, fault identification and capacity management more difficult.

Lack of standards is another issue in edge computing. While the cloud computing market has settled on a relatively small number of dominant platforms and application programming interfaces (APIs), the edge is "wild west" of platforms. A wide range of edge computing frameworks, communication protocols and hardware platforms makes it difficult to maintain interoperability and raises concerns over vendor lock-in. Security, as noted above, is one such challenge: the physical exposure of edge nodes to an open environment raises security issues that are not present with a data center.

B. Challenges in Cloud Computing

Although cloud computing is mature and popular, it has its flaws. Latency will continue to be an intractable issue for a purely centralized architecture: the speed of light constrains the physical minimum round-trip latency between a device and the data center, and this is limited but not overcome by further investment in network infrastructure. For the growing and critical set of latency-sensitive applications, the cloud-only platform is not an option no matter what else is done.

Data sovereignty and compliance is an emerging concern for global enterprises. Regulations governing the jurisdiction in which personal data can be stored and processed differ across jurisdictions, and the location of cloud data centers does not necessarily match customers' regulatory requirements. Vendor lock-in is also an issue: organizations with highly integrated services that have adopted the specific services of a given cloud provider may face significant challenges if they attempt to move to another cloud platform or repatriate workloads to onpremises data centers.

VII. FUTURE TRENDS

A. Hybrid Edge-Cloud Architectures

The future of the field is hybrid platforms that determine the placement of workloads across a continuum of computing assets - from resource-constrained end devices through intermediate edge devices to massive data centers. Instead of edge and cloud being seen as a method of choice, the general trend is to see them as part of a hierarchy of computing. Models for task offloading (algorithms that decide the most effective way to pull or push the computation to the next level of computing) are a hot topic [9].

Federated learning is an exciting use case for edge and cloud. In this approach, the models being trained on edge systems using locally collected data, and only model weights (not the training data) are shared with an aggregation server. This method ensures privacy of training data, eliminates the need to transfer potentially large quantities of data, and facilitates the creation of global models from remote data that cannot feasibly be professionally collected. With the development of federated learning and increasing capabilities of edge devices this paradigm is likely to be adopted in the coming years in healthcare, finance and telecom sectors.

B. 5G and Beyond

The rollout of fifth-generation (5G) telecommunication systems is likely to dramatically broaden the applicability of edge computing. 5G networks offer much higher speeds, lower delays and more devices per square kilometre than current networks, enabling both more powerful and more efficient deployments of edge computing. The MEC standardized framework for bringing edge computing closer to the network infrastructure in the radio access network of a cellular system is particularly important here. MEC makes it possible to deploy application workloads within the cellular network infrastructure, delivering latencies of 1-5 milliseconds sufficient for even the most stringent real-time requirements.

C. AI at the Edge

Continued miniaturization of neural network inference accelerators is allowing increasingly powerful artificial intelligence functions to be performed at the edge. Specialised edge AI chips, neural processing units integrated in mobile system-on-chip (SoC) designs, and new approach to model pruning and quantization algorithms are making it possible to perform inference with performance levels available only in data centres just a few years ago. This is poised to result in the use of complex perception, language processing and cognition capabilities being deployed on edge devices across many application domains, including agriculture, manufacturing, retail and logistics.

VIII. CONCLUSION

In this paper, we have provided an overview of edge computing and cloud computing, as well as a comparative analysis of their architectures, performance, security, scalability, and use cases. The study reveals that the two paradigms are not competing but rather complementary, each with its unique strengths that make it suitable for different use-cases and design constraints.

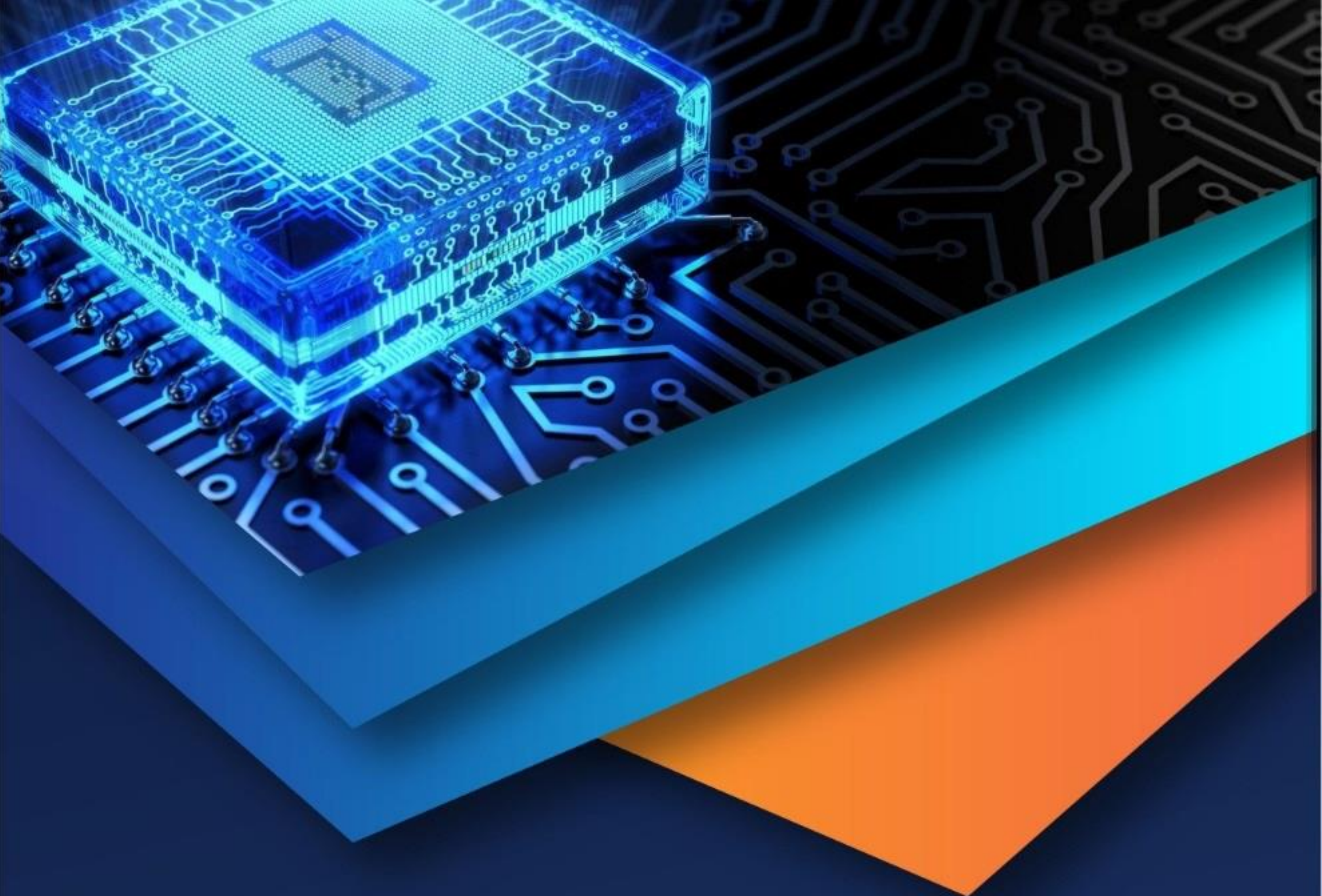
Use cases for edge computing are those applications that need low latency, location awareness and reduced load on the wireless network, and are ideal for any real-time application such as autonomous driving, process automation, IoT medical monitoring, and smart cities. Alternatively, cloud computing has unrivalled scalability and computational power, and a rich service ecosystem, making it best suited for global applications, complex workloads, and computational workloads such as large-scale machine learning training.

The findings discussed in this paper clearly indicate that the future of distributed computing will not be about prevailing turf of either paradigm; rather it will be about advanced hybrid computing systems that exploit the benefits of both edge and cloud. Smart workload placement, federated machine learning and the advancement of edge management tools will together make it possible for systems to have the edge and the cloud in one package. With the roll-out of 5G networks, which will increasingly deliver high-speed, low-latency connectivity in more settings, and as edge hardware capabilities improve, the edge-cloud distinction will increasingly blur.

Researchers need to develop standardised models for hybrid edge-cloud systems, extend the security mechanisms available in edge environments with limited resources, and build models of the cost of ownership for different deployment paradigms. We are at the level of transition and what researchers and practitioners do this time will determine where the computing world will be in future.

REFERENCES

[1] Statista Research Department, "Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2030," Statista, 2023.
[2] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, NIST Special Publication 800-145, Gaithersburg, MD, USA, 2011.
[3] M. Satyanarayana, P. Bahl, R. Caceres, and N. Davies, "The Case for VM-Based Cloudlets in Mobile Computing," IEEE Pervasive Computing, vol. 8, no. 4, pp. 14–23, Oct.–Dec. 2009.
[4] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, Oct. 2016.
[5] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Communications of the ACM, vol. 53, no. 4, pp. 50–58, Apr. 2010.
[6] R. Roman, J. Lopez, and M. Mambo, "Mobile Edge Computing, Fog et al.: A Survey and Analysis of Security Threats and Challenges," Future Generation Computer Systems, vol. 78, pp. 680–698, Jan. 2018.
[7] S. Subashini and V. Kavitha, "A Survey on Security Issues in Service Delivery Models of Cloud Computing," Journal of Network and Computer Applications, vol. 34, no. 1, pp. 1–11, Jan. 2011.
[8] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, and M. Imran, "The Role of Edge Computing in Internet of Things," IEEE Communications Magazine, vol. 56, no. 11, pp. 110–115, Nov. 2018.
[9] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog Computing and Its Role in the Internet of Things," in Proc. 1st ACM MCC Workshop on Mobile Cloud Computing, Helsinki, Finland, Aug. 2012, pp. 13–16.
[10] M. Pham, N. A. Le, and K. Kim, "A Review of Edge Computing in the IoT Ecosystem: Performance, Latency, and Security," Journal of Cloud Computing: Advances, Systems and Applications, vol. 10, no. 1, pp. 1–20, 2021.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY