

Latency Reduction in IoT Applications Through Edge Computing

¹Garima Sahu, ²Pawan Kumar

¹MCA 1st Semester, ²Assistant Professor

^{1,2}Amity University Chhattisgarh, India.

¹Surbhisahu9499@gmail.com, ²Pawancsa91@gmail.com

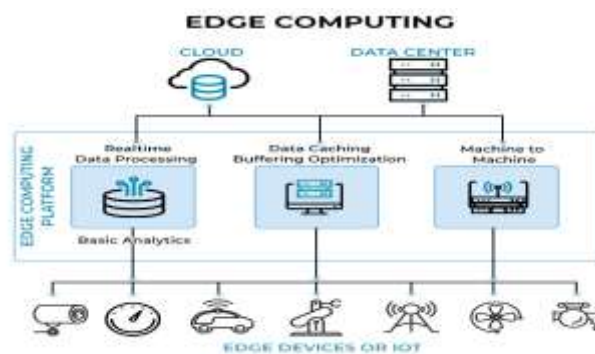
ABSTRACT

The rapid expansion of the Internet of Things ecosystem has created an urgent need for efficient data processing and analysis technologies. This review aims to systematically examine and compare edge computing, cloud computing, and hybrid architectures, focusing on their applications within IoT environments. The methodology involved a comprehensive search and analysis of peer-reviewed journals, conference proceedings, and industry reports, highlighting recent advancements. IoT has spawned numerous intelligent applications and services to benefit organizations, society, and consumer experiences. On the other hand, traditional computing methods are incapable of handling the demands of these services. The advent of cloud computing methods that provides software, platform, and infrastructure such as services have realized these applications. However, one of the critical obstacles of real-time cloud-based IoT applications is service response time. Edge computing solutions have been developed to address these issues. Results show that the reduced latency of edge servers bring significant benefits in terms of energy consumption. Experiments also show how the energy savings brought by edge computing are influenced by the prevalent direction of data transfer (upload vs download), load of the server, and daytime/nighttime operation. Therefore, edge-cloud-computing technology has recently evolved. This technology allows for data processing and storage at the edge of the network. This paper studies, in-depth, edge-computing architectures for IoT (ECAs-IoT), and then classifies them according to different factors such as data placement, orchestration services, security, and big data. Besides, the paper studies each architecture in depth and compares them according to various features. Additionally, ECAs-IoT is mapped according to two existing IoT layered models, which helps in identifying the capabilities, features, and gaps of every architecture. Moreover, the paper presents the most important limitations of existing ECAs-IoT and recomme Edge computing is transforming the Internet of Things (IoT) by enabling real-time data processing at the network's edge, reducing latency and enhancing security for smart systems.

Keywords: IoT, Edge computing, Latency reduction, Real-time processing, Edge devices, Cloud computing.

1. Introduction

The Internet of Things (IoT) is expanding into different aspects of our lives with technologies and applications in, for example, smart cities, healthcare, and smart homes [1]. Tens of billions of objects are connected to the Internet [2], and the industry expects 50 billion IoT devices to existing by 2020 [3]. However, IoT devices are limited in resources such as storage and processing power, which impacts the performance, security, reliability, and privacy of IoT-based solutions and applications [4,5]. Many applications are enhanced by integrating the IoT and cloud computing. Examples of such applications are in healthcare [6,7,8], smart cities [8,9,10], smart homes [11], smart metering [5], video surveillance [5] such as smart urban surveillance applications [12], agriculture [13], such as greenhouse environment-monitoring system [14], and smart mobility [15], such as smart tourism destinations [16]. Although IoT devices are limited in resources, cloud computing helps IoT in addressing such limitations [5,17]. IoT data are transmitted continuously from applications to a central storage unit, which is usually located in a cloud center [18]. Some IoT applications require low latency time and they may need real-time processing. Handling such requirements by cloud computing is not suitable [19]. Thus, edge computing is crucial for fulfilling these requirements by deploying cloud-computing-like capabilities at the edge of the network [20]. This survey paper focuses on current edge-computing architectures (ECAs) for IoT applications (ECAs-IoT). This paper introduces the ECA-IoT concept and surveys current ECAs-IoT and possible research opportunities. Among the promising features of Edge computing are included mobility support, location awareness, ultra-low latency, and proximity to the user [12]. These features make Edge computing suitable for different future applications like industrial automation, virtual reality, real-time traffic monitoring, smart home, smart sea monitoring and data analytics as shown in Fig. 1. Edge devices like routers, access points, and base station host different services (e.g., QoS, VPN, and Voice over IP etc.) [13]. These Edge devices act as a bridge that connects the smart mobile devices with the cloud. Several surveys (S. Yi et al. [14], L. M. Vaquero et al. [15], I. Stojmenovic [16], F. Bonomi et al. [17], T.H. Luan et al. [18], E. Ahmed et al. [19]) have studied various aspects of Edge computing like Fog computing while a few studies focus on the Edge computing domain like Mobile Edge computing that focuses on specific application domains. However, there is no comprehensive study has been completed yet that includes all aspects of Edge computing including Mobile-Edge, Fog and Cloudlet computing.



Literature review

Real-Time Processing in Edge-Based IoT Systems

Real-time processing is a critical requirement for many IoT applications that demand instantaneous data analysis and decision-making. Edge computing plays a pivotal role in meeting this requirement by reducing latency and enabling faster responses. This section reviews studies on real-time applications in edge-based IoT systems across key domains.

Autonomous Vehicles

Autonomous vehicles rely heavily on real-time decision-making for navigation, object detection, and collision avoidance. The integration of edge nodes with machine learning models for real-time image recognition further enhances the system's ability to detect pedestrians and other vehicles with minimal delay.

Moreover, Liu et al. (2021) proposed a hierarchical edge-cloud architecture for autonomous driving, where time-sensitive tasks such as obstacle detection are handled at the edge, while non-urgent tasks like long-term route optimization are offloaded to the cloud. This hybrid approach balances latency reduction with resource efficiency, making it a practical solution for large-scale deployments.



Healthcare

In healthcare, real-time processing is vital for monitoring patients, diagnosing conditions, and providing timely interventions. Wearable IoT devices, such as smartwatches and health trackers, leverage edge computing to perform localized data analysis, significantly reducing the latency involved in transmitting data to centralized servers.

For instance, Kumar et al. (2019) explored an edge-enabled healthcare monitoring system where wearable sensors transmit patient data to nearby edge nodes for processing. Edge computing also enhances telemedicine applications, where real-time video processing and

diagnostics are critical. Studies show that edge-based systems reduce jitter and latency during teleconsultations, improving the quality of remote healthcare services.



Industrial Automation

In industrial automation, real-time monitoring and control are crucial for maintaining operational efficiency and safety. Edge computing is extensively utilized to process data from IoT sensors in factories, enabling real-time detection of anomalies and predictive maintenance.

A study by Singh et al. (2021) introduced an edge-based system for monitoring industrial equipment. By analyzing sensor data locally Furthermore, real-time edge analytics helped optimize production processes by identifying bottlenecks and adjusting operations dynamically.



Smart Cities

Smart city applications, such as traffic management, public safety, and environmental monitoring, benefit greatly from real-time data processing. A case study by Lee et al. (2020)

on smart traffic lights showed that edge computing reduced response times in traffic signal adjustments, leading to a 30% improvement in traffic flow during peak hours.

These examples highlight the transformative potential of edge computing in enabling real-time processing for diverse IoT applications. By addressing the latency challenges of traditional cloud systems, edge computing has become indispensable in domains requiring immediate decision-making and responses. Future work in this area could further optimize edge architectures, integrate advanced AI techniques, and explore novel applications in emerging fields.

Importance of Latency in IoT Applications

Real-time Latency Reduction

Many IoT applications are essentially **advanced monitoring systems**: they collect data, analyse the data and then trigger actions based on the insights generated. In some cases, this is done on an hourly or daily basis, or only when triggered by a specific interaction with the device. Edge computing can provide a benefit to IoT when these insights are needed in real-time. By providing compute closer to the IoT device, data collection and analytics take place at a physically closer location (i.e. often within the same country or region, perhaps even on the premises, rather than in a large centralised data centre). In doing this, **network latency** is reduced as the round trip to the data centre and back is shorter. In this way, edge computing can optimise IoT applications that require real-time actions (e.g. cooling systems turned on as soon as a sensitive piece of machinery starts to overheat).

Enhancing IoT Security

A major headache for many with the IoT ecosystem is how to manage security as more and more devices are connected. Malware can be used to harness IoT devices to perform DDoS attacks, for example. While edge computing is unlikely to be, in and of itself, more secure than a private cloud, it does have the benefits of being more local. For companies concerned about storing data in locations which, for example, have different data protection laws than where the data is being generated, edge computing can provide some security benefits. Particularly if the edge servers are located on the premises, companies can be sure that data never leaves their own local perimeter and can control all access to the servers storing the information.

IoT edge platforms: the specific requirements of IoT on edge computing

In general, operators are looking to roll out edge computing sites in their network using COTS (common off the shelf) servers without specific hardware for specific applications. While today on-premise edge deployments are generally very bespoke and are deployed and integrated for one specific customer, in the future these deployments will also have to become more standardised.

There are a few IoT-specific requirements that edge providers will have to manage. For example, some IoT applications which require heavy duty **AI or data analytics at the edge**

may have specific hardware requirements, such as servers with GPU capacity. Edge gateways supporting IoT devices will also have to support connectivity for multiple types of device communication such as ZigBee and Bluetooth, as well as connectivity via cellular and Wi-Fi technology.

In addition, operator need to consider the type of IoT platform they are using and how it supports and integrates with their edge infrastructure. Many operators with IoT businesses rely on third party vendor platforms such as PTC ThingWorx and Software AG's Cumulocity to manage their IoT networks. As demand for edge applications continues to grow, platform vendors are increasingly investing in introducing edge-specific capabilities to their platforms.

Challenges and Limitations

Despite its advantages, edge computing faces several challenges:

Resource constraints: Limited processing power and storage compared to centralized cloud systems.

Interoperability: Diverse IoT devices and platforms require standardization for seamless integration.

Security: Edge devices are susceptible to physical and cyber threats due to their distributed nature.

Future Research Directions

To enhance latency reduction in IoT systems through edge computing, future research could focus on:

AI-driven optimization: Developing intelligent algorithms for load balancing and task scheduling at the edge.

Energy efficiency: Designing low-power edge devices for sustainable IoT deployments.

5G integration: Leveraging 5G networks for ultra-low-latency IoT applications.

Blockchain at the edge: Ensuring secure and transparent transactions in decentralized systems.

Conclusion

The rapidly growing IoT ecosystem demands robust and efficient solutions for real-time data processing and security management. Edge computing has emerged as a transformative technology, addressing critical requirements for instantaneous data analysis and decision-making. By processing data closer to the source, edge computing

reduces latency, ensuring faster responses for time-sensitive IoT applications across domains such as smart cities, healthcare, and industrial automation.

Moreover, edge computing offers distinct advantages in security and data sovereignty. While it may not inherently provide greater security than private cloud infrastructures, its local nature allows organizations to mitigate risks associated with cross-border data transfers and varying data protection laws. By keeping sensitive information within on-premises edge servers, companies retain control over their data, enhancing both security and compliance.

In conclusion, edge computing's ability to balance real-time performance and localized security makes it a pivotal component of IoT architectures. However, to fully unlock its potential, future advancements should focus on addressing its inherent limitations, such as scalability and resource constraints, and fostering seamless integration with cloud and hybrid systems for more versatile IoT solutions.

References

1. Salman, O, Elhaji, I., Kayssi, A., & Chehab, A. (2015). Edge computing enabling the Internet of Things. In *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)* (pp. 603–608). IEEE.
2. Gupta, N., Bafna, K., & Gupta, N. (2011). Efficient data replication algorithm for mobile Ad-hoc networks. In *International Conference on Information and Network Technology* (pp. 164–170).
3. Hara, T., & Madria, S. K. (2009). Consistency management strategies for data replication in mobile ad hoc networks. *IEEE Transactions on Mobile Computing*, 8(7), 950–967.
4. Sasaki, Y., Hara, T., Ishikawa, Y. (2019). Efficient framework for processing top-k queries with replication in mobile ad hoc networks. *GeoInformatica*, 23(4), 591–620.
5. Sethi, P., & Sarangi, S. R. (2017). Internet of things: Architectures, protocols, and applications. *Journal of Electrical and Computer Engineering*, 2017(1), 1–25.