Scalable Proximity Cloud Architecture for Enhancing IoT Service Performance

Anil Rajput

Professor Computer Science & Mathematics,

Govt. C. S. A. P. G. College, Sehore, MP

e-mail:dranilrajput@hotmail.com

Kshmasheel Mishra

Former Reader Computer Science,

Vikram University, Ujjain, MP

e-mail:mishra_ks@hotmail.com

Abstract

This paper presents a novel Proximity Cloud Computing Architecture designed to enhance the performance of IoT devices by positioning computational resources closer to the network edge. This proximity reduces latency, enabling real-time data management and analytics, which is crucial for IoT applications requiring immediate response, such as healthcare monitoring, autonomous vehicles, and industrial automation. By processing data locally, the architecture optimizes bandwidth usage, minimizes network congestion, and enhances system responsiveness. Additionally, keeping sensitive data closer to its source improves security. The proposed architecture serves as a key enabler for the next generation of IoT, fostering smarter, faster, and more secure operations across various industries.

Keywords

IoT Services, Latency Reduction, Edge Computing, Micro Data Centers, IoT Data Security, Cloud Computing, Autonomous Vehicles, Healthcare Monitoring, Industrial Automation, Software-Defined Networks (SDN), Scalable Architecture.

1.0 Introduction to Proximity Cloud Computing and IOT

Proximity Cloud Computing represents an innovative evolution in cloud technology by extending computational and storage resources to the network edge, closer to the data source. This approach addresses the limitations of traditional cloud models by decentralizing computation, storage, and networking services. By positioning these resources nearer to the Internet of Things (IoT) devices, proximity cloud computing significantly reduces latency, enhances real-time data handling, and improves overall system efficiency. This is particularly beneficial for applications requiring prompt data analysis and response, such as those in autonomous vehicles, smart cities, and healthcare monitoring systems [1].

In the realm of IoT, which connects various IP-enabled objects like sensors, vehicles, and smart home devices, the traditional cloud model is extended to encompass edge computing. This shift is crucial for managing the vast amounts of data generated by IoT devices [2]. The integration of microservices into edge devices or local servers is essential for efficiently processing and storing data in proximity to its source. This not only optimizes resource usage but also alleviates the load on centralized cloud infrastructures, contributing to more resilient and scalable systems.

Despite the advancements, the field of Proximity Cloud Computing remains an evolving research area with considerable gaps, particularly in applications and platforms. The complexity of modeling these applications arises from the need to aggregate data from diverse IoT devices that use various protocols and standards. Creating adaptable and easily deployable applications remains a significant challenge [3].

Resource management is a critical aspect of cloud computing, especially in IoT environments where resources are varied and dynamic due to the diverse capabilities of different devices. This research aims to develop a more refined approach to resource provisioning within proximity cloud architectures. The core objective is to improve service performance through enhanced resource management strategies [4].

The research addresses two primary questions:

RQ1: How can new and improved resource provisioning approaches be designed for proximity cloud computing?

RQ2: Do these innovations offer substantial improvements over existing cloud computing environments?

By focusing on these questions, this study proposes strategies for better resource placement and management, ultimately enhancing the efficiency and effectiveness of IoT services.

2.0 Review of Related work

2.1 Latency Management in Cloud Computing

Several works are available to handle the latency issues in Cloud Computing. IoT typically leverages mobile computing networks to manage the rapidly growing mobile data traffic. Radio access points can function as an infrastructure for a local mobile cloud computing platform. Job image replicas can be localized and stored on local server, and since multiple clients may share a storage server for the same task, the location of the stored task image significantly impacts the performance of the service and quality of service parameters (QoS parameters). This is especially true for the task completion time which is certainly better in such localized setups [5].

A job can be administered on a nearby computation facility instead of solely computing it on the thin-client device. Additionally, transmission latency is a significant factor that cannot be overlooked. The authors concentrated on determining the finest task-image delegation and computation strategy and decision decentralization to bring down the job completion time and improve the throughput of the system [6].

2.2 Software Defined Networks Management in Cloud for IOT

The fundamental concept of Software Defined Networks is the separationamongst network control (the control plane) and data forwarding functions. The controller has the responsibility of functioning as the overall decision maker for the control logic. It also performs the role of the network director and is responsible for directing the globalperformance of the network [7].

This innovative networking paradigm provides numerous advantages over conventionalnetworking functionalities and add new dimensions to the methods of controlling the flow of packets in the network. This also reduces the need of independently accessing the hardware and configure multiple devices disseminatedall over the network [8].

SDN-based architectures offer a centralized controller with a comprehensive understanding of the network state, capable of managing network organizationautonomouslywithout concern for vendor specific policies. This enables direct control, programming, orchestration, and management of network controller, thereby saving significant workforce and resources [9].

2.3 Computation Management in Micro Datacenter

In several IOT sensor networks the time component is critical and the dispatch of sensor data to cloud may induce unpredictable delays. Effective distribution of computation tasks with timely resource scheduling and management enable data centers to respond appropriately to various situations, enhancing customer satisfaction [10].

3.0 Proposed Proximity Cloud Architecture

The proximity cloud architecture evolves to an advanced version of cloud computing by putting more prominence on the calculation and storing of data close to the network. The proximity cloud architecture progresses cloud computing by betterfocus on the calculation and storing of data at the boundary of the sensor network of IOT devices. This innovative approach leverages the proximity of IOT devices to end-users, allowing for faster data processing and reduced latency compared to traditional cloud models. As a result, proximity cloud computing provides a robust and efficient solution for managing the increasing demands of modern, data-intensive applications.

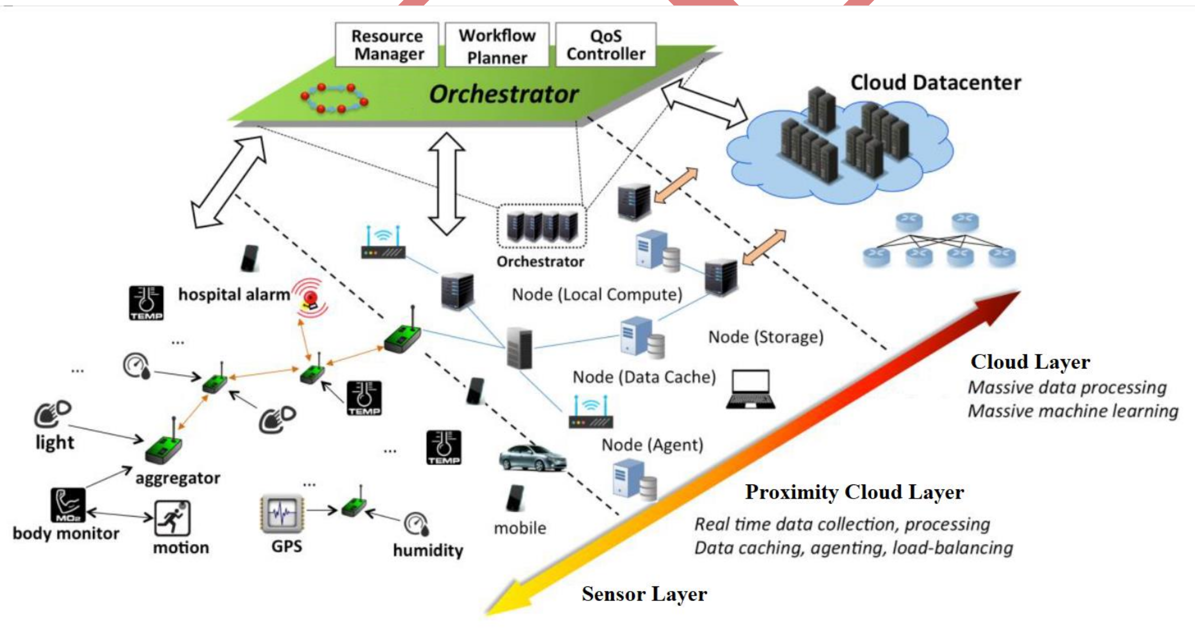

Figure 1: Proposed Proximity Cloud Architecture for IOT

This proximity cloud can be envisaged as comprising of three distinct blocks of participating units:

(a) Block 1: The lowest layer, comprising all IoT devices,

(b) Block 2: The central layer, referred to as the Proximity Cloud Computing layer is located nearby or sandwiched between the fundamental IOT nodes and the far-off cloud layer

(c) Block 3: The topmost layer, known as the Cloud Computing layer. This is the datacenter which is responsible for all high-end computing needs. We intend to reduce the load of this layer and share the computation work with the central layer.

The orchestration layer is a highly scattered layer which is as dispersed as the fundamental infrastructure and services.

The key technologies and components to achieve this include:

- A software agent that is not very complex or powerful but is capable of handling orchestration functionality. This agent is also responsible for management of system and performance requirements, which can be embedded in various proximity devices.
- Dispersed, persistent storage capable of storing policies and source metadata (such as capability and performance) that supports large scale transactions updates and retrievals.
- A messaging and communication bus that can support scalable cloud operations.
- A dispersed policy repositorythat supports an integrated global versioningalong with localized administrationabilities.

4.0 Service stack of proposed model

The proposed proximity cloud comprises of several services. Some of the prime service components are:

- **Flow and Task Placement**
- **Knowledge Base**
- **Performance Prediction**
- **Raw-Data Management**
- **Monitoring**
- **Profiling**
- **Resource Provisioning**
- **Security**

This service stack is tentative and may be up scaled or down casted depending on individual architecture.

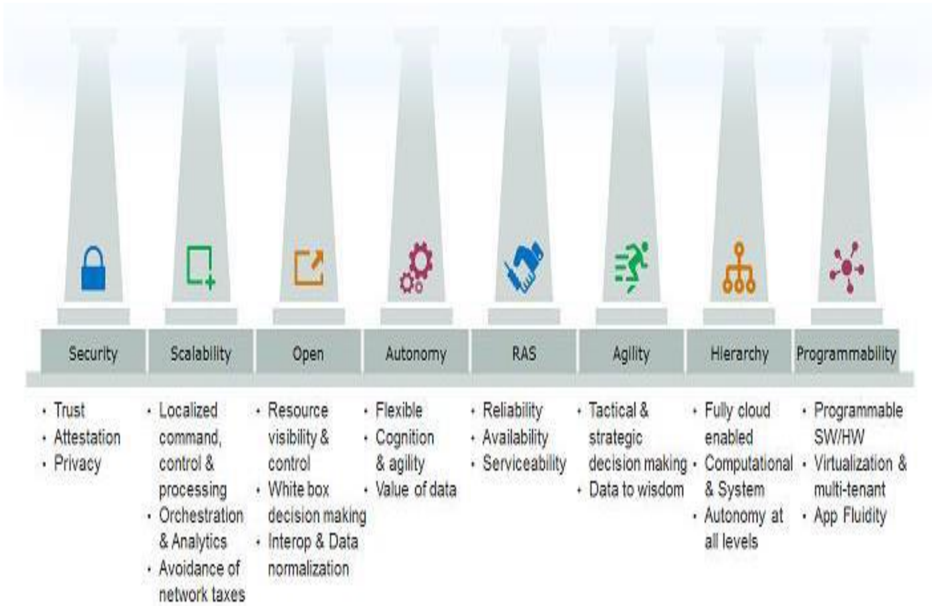

Figure: Service Stack of Proximity CloudComputing

5.0 Significance and Advantages of Proximity Cloud Computing

Proximity cloud computing offers scalable, on-demand storage and processing services tailored to meet IoT requirements. For applications such as health IOTdevice networks and emergency response that demand fast action and low latency, minimizing the latency on account of transmitting data to and from distant cloud servers is crucial. Sending large volumes of data to the cloud for storage and computation can strain network bandwidth and prove inefficient.

To tackle these challenges, proximity cloud computing advocates utilizing nearby computing resources for local storage and initial data processing near IoT sensors. This approach aims to alleviate network congestion, expedite data analysis, and enhance decision-making speed. The key innovations of this proposed architecture include:

1. Performance Improvement:

Achieving low latency is a primary motivator for adopting proximity cloud architectures. Various stakeholders have specific requirements and design considerations to ensure this goal is met. These considerations encompass time-critical computing, time-sensitive networking, and network time protocols, among others. Addressing latency is a pervasive concern due to its impacts on system performance and deployment scenarios. The architecture benefits from efficiently pooling local idle resources from participating IOT devices dynamically.

2. Improved Security:

Security is indispensable for the acceptance of proximity cloud deployments. Secure underlying silicon alone isn't sufficient; security vulnerabilities in upper-layer software or hardware compromise overall security. Ensuring data integrity is crucial, especially for devices lacking adequate security measures, guarding against both deliberate and accidental data corruption. The localized approach provides supplementary security layers to guarantee safe and trusted transactions.

3. Manageability:

Effectively managing all aspects of proximity cloud deployments, including reliability, availability, and serviceability (RAS), as well as integrating DevOps practices, is critical across all computing hierarchy layers. Improved management includes better cognitive understanding of client-centric objectives, enabling greater autonomy in operations.

4. Data Analytics and Control:

Proximity cloud nodes need autonomous capabilities, necessitating that data analytics is carried out locally, but this again is to be paired with policy based operative control mechanisms. The policy driven control operations should occur at the suitablenode within the hierarchy. This automation and control of the computing node depends on the situation requirements. Control agility, rapid innovation, and scalable affordability are enhanced under a unified infrastructure, supporting low latency.

6.0 Conclusion

Proximity cloud computing addresses performance challenges in advanced digital IoT deployments, focusing on managing system efficiency, reducing the latency in response by local computing, and improving network efficiency by reducing unnecessary load on the network. It's crucial to understand that cloud and proximity cloud computing complement each other on a reciprocallyadvantageous continuum. Proximity cloud computing doesn't replace traditional cloud solutions but collaborates with them to fulfill the specific needs of selected use cases. Some functions are inherently more suitable for execution in proximity cloud nodes, while others benefit from traditional cloud environments. As proximity cloud computing evolves, traditional backend cloud systems will remain a substantial part of computing infrastructure.

References

[1] H. Niedermayer, R. Holz, M.-O. Pahl and G. Carle, "On using home networks and cloud computing for a future internet of things," in *Future Internet-FIS 2009: Second Future Internet Symposium*, FIS 2009, Berlin, Germany, September 1-3, 2009.

[2] G. M. Lee and N. Crespi, "Shaping future service environments with the cloud and internet of things: networking challenges and service evolution," in *Leveraging Applications of Formal Methods, Verification, and Validation: 4th International Symposium on Leveraging Applications, ISoLA 2010, Proceedings, Part I 4, pp. 399-410. Springer Berlin Heidelberg*, Heraklion, Crete, Greece, October 18-21, 2010.

[3] L. I. Bo-Hu, Z. Lin, W. Shi-Long, T. Fei, C. A. O. Jun-Wei, J. Xiao-dan, S. Xiao and C. Xu-dong, "Cloud manufacturing: a new service-oriented networked manufacturing model," *Computer integrated manufacturing system* , vol. 16, no. no. 01 (2010), 2010.

[4] M. Wu, T.-J. Lu, F.-Y. Ling, J. Sun and H.-Y. Du., "Research on the architecture of Internet of Things," in *3rd international conference on advanced computer theory and engineering (ICACTE), vol. 5, pp. V5-484. IEEE, 2010.*, Honkong, 2010.

[5] Z. Wan, "Cloud Computing infrastructure for latency sensitive applications.," *IEEE 12th International Conference on Communication Technology,* , pp. pp. 1399-1402, 2010.

[6] S. K. Barker and P. Shenoy, "Empirical evaluation of latency-sensitive application performance in the cloud," in *Proceedings of the first annual ACM SIGMM conference on Multimedia systems, pp. 35-46. 2010.*, Singapore, 2010.

[7] P. Pavol, H. Pavol, T. Halagan and I. Drozd, "Future Networks-SDN & NFV," *SIGCOMM Computer Communication Review,* vol. 38, no. 2, pp. 69-74., 2008.

[8] R. Buyya, C. S. Y. S. Venugopal, J. Broberg and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation computer systems* , vol. 25, no. 6, pp. 599-616, 2009.

[9] D. Hilley, "Cloud computing: A taxonomy of platform and infrastructure-level offerings," in *Georgia Institute of Technology, Tech. Rep (2009): 44-45.*, 2009.

[10] K. A. G. G. a. J. R. H. Church, "On Delivering Embarrassingly Distributed Cloud Services," *HotNets,*, vol. 3, no. 6, pp. pp. 55-60., 2008.