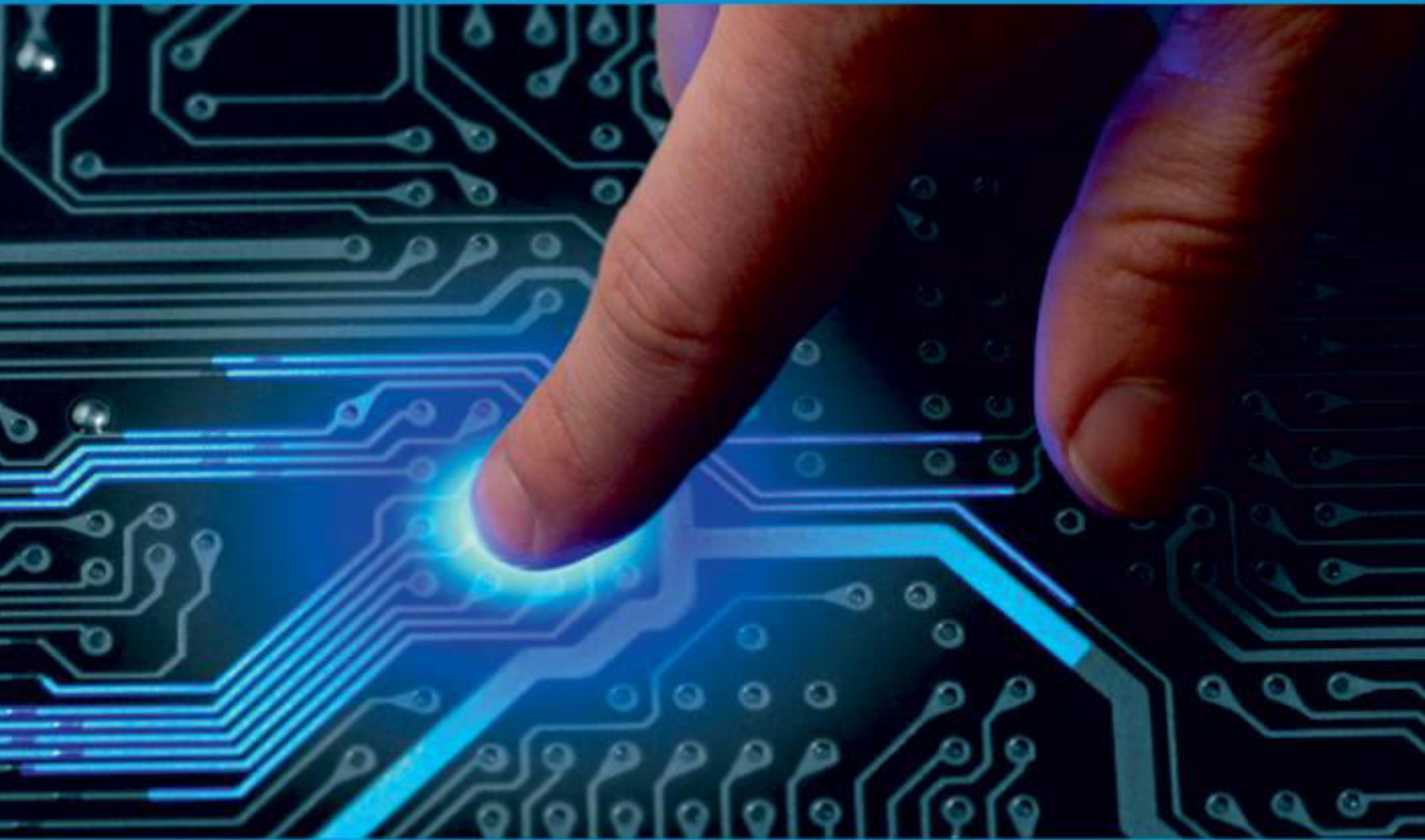
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211033

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

LBRO: Load Balancing for Resource Optimization in Edge Computing

Mala K, Vinay T P, Bhavana Y, Chandana K R

Assistant Professor, Department of Information Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

Assistant Professor, Department of Information Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

UG Student, Department of Information Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

UG Student, Department of Information Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

ABSTRACT: Solutions based on mobile cloud computing and edge computing offer ways to delegate work to mobile devices with low resources. While edge computing offers closer proximity-based solutions, mobile cloud computing offers cloud solutions that are located farther away. Due of distance and reliance on the Internet, remote cloud solutions face issues with network latency and restricted capacity. However, because the edge node is accessible within the same network, edge-based solutions address these issues. In light of future information and communication technology infrastructure, the adoption of Internet of Things-based solutions is growing, which leads to the enormous expansion of digital equipment and an increase in the load on edge devices. To prevent resource congestion, a load balancing system at the edge level is therefore necessary. The user's preferences about edge resources, like PCs or mobile devices, must be taken into account for load balancing at the edge. To prevent overprovisioning vital resources, a user must specify which resources can be saved for other devices. In order to handle load balancing issues in edge computing while taking user preferences into account, we introduce Load Balancing for Resource Optimization (LBRO), a collaborative cloudlet platform. A comparison between the suggested method and the traditional edge-based method reveals that the suggested method offers noticeably better CPU, memory, and disk utilization outcomes.

KEYWORDS: Mobile cloud computing, mobile edge computing, cloudlet computing, Internet of things, cloud federation.

I. INTRODUCTION

Mobile devices have restricted resources, such as a central processing unit (CPU), memory, energy, and network capabilities. With the growth of resource demanding applications, these limitations become more pronounced, intensifying the resource constraint issue on mobile devices. Cloud computing provides a resource-abundant environment for these devices, enabling resource sharing and load balancing. Virtualization allows the resources of a physical machine to be shared effectively. Cloud computing offers several service-oriented architectures, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). This paper focuses on the IaaS model, where a task is packaged as a virtual machine (VM) and deployed on a physical server (cloudlet). The mobile cloud computing (MCC) model is used to offload compute-intensive tasks from mobile devices to the cloud, alleviating resource limitations. In the MCC model, mobile devices communicate directly with remote servers over wireless Internet.

However, challenges such as latency, limited bandwidth, and the need for consistent connectivity limit the effectiveness of this approach. Edge-based solutions, such as mobile edge computing (MEC), fog computing, and cloudlet computing, offer proximity-based alternatives that help overcome these limitations. Among these, cloudlet-based solutions are more versatile, providing extensive computational resources, diverse features, and higher bandwidth without relying on specialized equipment, unlike MEC and fog computing. A cloudlet acts as a mini-cloud with substantial computing power and a stable Internet connection within the local area network (LAN) to serve nearby devices. Cloudlet-based solutions are seen as particularly effective for Internet of Things (IoT) applications and smart cities due to their quicker response times compared to mobile cloud computing (MCC). However, as more devices connect to cloudlets, the workload often surpasses their capacity. In such cases, cloudlets forward the excess requests to

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

a remote cloud to handle the load, effectively reverting to a traditional MCC model and diminishing the benefits of edge computing

II. LITERATURE SURVEY

This section offers information about innovative edge computing methods. The relevant literature is examined from the standpoint of resource optimization and load balancing in an edge federated context.

A queuing network approach has been introduced using a multi-edge and user-based model, which efficiently manages user-to-edge device mapping and edge device placement. This approach targets mobile users who are constantly moving and focuses on a Metropolitan Area Network (MAN). The scheme incorporates load balancing by assigning virtual machines (VMs) to edge devices with sufficient resources. Its primary focus is on minimizing total migration time rather than download time, achieved by adapting to Wide Area Network (WAN) bandwidth and the load on edge devices. The VM state on the target edge device is compared with the source to calculate and synchronize any differences before shutting down the source VM. Delta encoding is applied to calculate these differences, which are then de-duplicated and compressed prior to transfer. To reduce latency, the user VM is positioned closer to the source edge device. An SDN-based solution, called Mobi Scud, establishes a mini cloud in the core of the Radio Access Network (RAN) to host users' VMs. These VMs help users perform compute-intensive tasks and handle control messages from mobile devices. Mobi Scud monitors these VMs, ensuring they stay close to users as they move, optimizing the migration process with minimal service disruption. Additionally, since users often switch to Wi Fi indoors, RAN services need to adjust accordingly. An ad hoc scheme has been proposed that enables peer devices to share and borrow resources.

Devices in communication are designated as master and slave, where the slave device offers free resources, and the master device borrows them. This collaborative approach works well for smaller networks, assuming uninterrupted task offloading. Another ad hoc scheme utilizes short-range radio technology to create a peer-to-peer (P2P) network among mobile devices, where participants are divided into two roles: computational service providers with sufficient resources and clients who need resources. An opportunistic approach helps devices find suitable peers for lending services, making this scheme effective for smaller networks with brief service needs. Another scheme, DRAP, introduces middleware between mobile and edge devices. Devices with available resources form a group and act as edge devices, which other mobile devices can access. DRAP's operations include resource discovery, unused resource calculation, and management of device roles.

It is highly adaptable, allowing devices to reconfigure dynamically as nodes join or leave. For reliability, some buddy nodes maintain logs to support service continuity in case of failures, and an incentive-based system encourages user participation. In another scheme, a combination of mobile devices, edge devices, and remote cloud is used for task offloading. Mobile devices in need of resources can connect with nearby edge devices or a remote cloud. Offloading to a nearby edge device reduces latency and avoids internet bandwidth usage. If offloading to the remote cloud is required, however, the model becomes a standard Mobile Cloud Computing (MCC) Model, with higher latency due to internet dependency. An edge-based scheme further utilizes predefined VM templates to meet user requirements on edge devices. It selects the best-matching template and applies infrastructure-level customizations, which are rolled back post-task to maintain infrastructure consistency. VMs isolate infrastructure changes from those at the guest OS level, although workload sharing and load balancing among tenant VMs are not discussed. A peer-to-peer (P2P) scheme is proposed to address the issue of selfish behavior among participating devices through a point-based incentive model. In this system, there are two types of devices involved in the collaboration. Devices can earn points by sharing their resources with others, while devices needing resources can spend points to obtain them. This approach also incorporates a social responsibility framework within a community group based on a pre-trust model, ensuring that only trusted devices participate in the collaboration.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

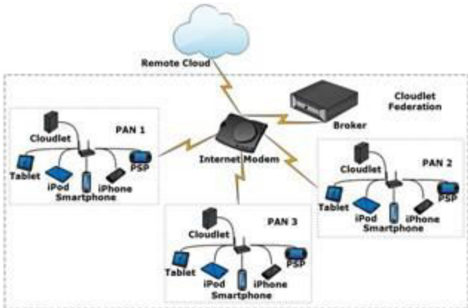
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. METHODOLOGY

This section contains the design of the proposed collaborative model and its architectural details. In addition, the proposed algorithms are discussed in detail.

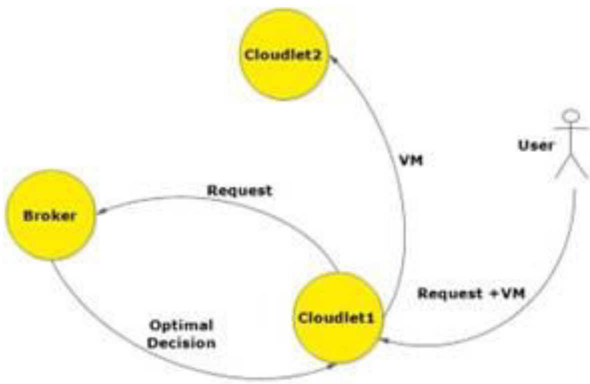
COLLABORATIVE MODEL

The proposed federated cloudlet model, illustrated in Figure 1, addresses resource shortages at cloudlets and minimizes the need for forwarding requests to the remote cloud. Cloudlets within this model may have different owners, administrative domains, and affiliations with various Cloud Service Providers (CSPs). Each cloudlet owner sets resource preferences through a federation broker, locking only the preferred resources based on priority and allocated percentage. This approach ensures that the owner's tasks and services on the cloudlet node are not impacted by overprovisioning, offering reassurance regarding performance stability.



LOAD-AWARERESOURCEALLOCATION

Since all the applications are placed on a cloudlet shared resource based on time thus an increased wait time is observed for the new applications and ultimately the poor resource availability degrades the performance of the over-all system and all applications running on it. The resource limitation not only affects the performance but also forces the cloudlet to forward the requests to the remote cloud to manage the load. The current implementations of cloudlets are only focused on addressing the distance, limited bandwidth, and latency challenges considering only a standalone cloudlet or group of cloudlets at the same location shared via LAN. Our pro-posed approach offers a load-aware collaborative scenario in which cloudlets share the user-preferred load and resources with peer cloudlets, managed by a centralized broker in the federation that may extend to a MAN or WAN.



CALCULATION OF RESOURCE UTILIZATION

The problem of finalizing the optimal cloudlet is challenging due to dependency on multiple variables including resources such as CPU, memory, storage, and bandwidth. Weights are assigned to each resource by the owner of the edge device to segregate spare resources for sharing. For example, an owner 'x' wishes to spare 20% of the CPU, 30% of the memory, 10% of his disk storage, and 5% of the bandwidth to take part in the sharing process for some other user

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

'y' to acquire these resources for the execution of some task. The owner can simply assign weights 2, 3, 1, and 5, respectively to each available resource for sharing.

IV. EXPERIMENTAL RESULTS

CPU Utilization

The LBRO system demonstrates more stable CPU utilization compared to conventional models. When subjected to peak load, the LBRO model, with its load balancing mechanism, maintains CPU utilization at optimal levels by distributing tasks across multiple cloudlets. This contrasts with the conventional approach, where a single cloudlet might experience spikes and drops in CPU load, leading to inefficiency.
Reference: "The results of the proposed model show a stable CPU utilization as compared to the conventional model..."

Memory Utilization

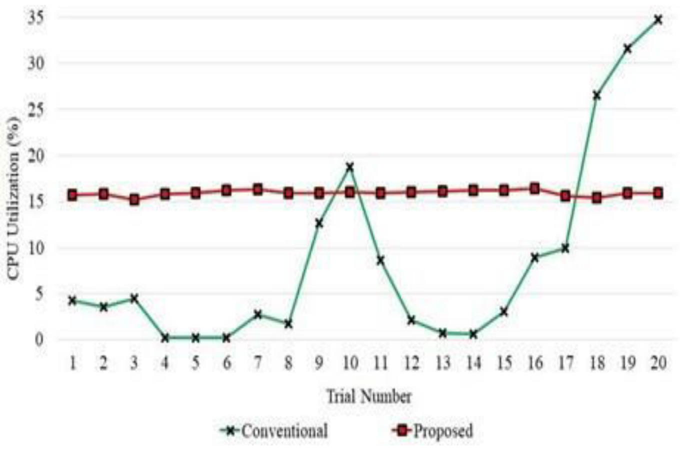
LBRO's collaborative load balancing minimizes memory bottlenecks. As observed, memory usage in LBRO models stabilizes, preventing the system from reaching critical memory states. This contrasts with conventional models that may experience complete memory usage, leading to potential performance degradation.
Reference: "The results clearly show an elevated level of memory utilization by the conventional model completely utilizing the memory causing performance degradation"

Disk Storage Utilization

LBRO's distributed task approach reduces disk utilization pressures by balancing data storage requirements. Unlike traditional models where disk operations peak and affect performance, LBRO limits disk activity by reducing the number of read/write operations.
Reference: "The results clearly show an elevated level of disk utilization by the conventional model..."

Network Latency

LBRO's optimal resource allocation minimizes network latency by selecting cloudlets with available resources close to the task origin. This leads to faster response times compared to the conventional setup, where tasks are forwarded to remote clouds if resources are unavailable locally.
Reference: "The broker...dispatches the request to the optimal cloudlet having adequate resources with minimum latency"



Summary

The LBRO approach provides substantial improvements over traditional models in terms of resource utilization and system stability. By using collaborative load balancing, LBRO ensures each cloudlet operates within optimal parameters, enhancing both performance and scalability in edge computing environments.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

The LBRO model for load balancing and resource optimization in edge computing successfully addresses key limitations of conventional cloudlet and mobile cloud computing models. By implementing a collaborative cloudlet approach that considers user-defined resource preferences, LBRO significantly reduces CPU, memory, and disk utilization spikes, leading to more stable and efficient system performance. The approach minimizes latency by keeping tasks closer to users, as the broker selects optimal cloudlets with available resources and low response times. This strategy not only improves resource utilization but also enhances scalability by leveraging a federated network of cloudlets across metropolitan and wide area networks.

Compared to traditional cloudlet models, LBRO prevents resource overload by managing requests based on real-time resource availability and migrating tasks to underutilized cloudlets as needed. This flexibility allows the model to avoid forwarding requests to remote clouds, which typically introduces higher latency and bandwidth issues, effectively preserving the benefits of edge computing.

In future work, developing a commercial-grade platform for LBRO could further optimize edge resource allocation and extend capabilities to include dynamic weight allocation for resources based on machine learning models, ultimately enabling even more adaptive and efficient resource sharing.

REFERENCES

[1] M. Z. Nayyer, I. Raza, and S. A. Hussain, “Revisiting VM performance and optimization challenges for big data,” in *Advances in Computers*, vol. 114, Amsterdam, The Netherlands: Elsevier, 2019, pp. 71–112.
[2] W. Liu, W. Gong, W. Du, and C. Zou, “Computation offloading strategy for multi user mobile data streaming applications,” in *Proc. 19th Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2017, pp. 111–120.
[3] M. Z. Nayyer, I. Raza, and S. A. Hussain, “A survey of cloudlet-based mobile augmentation approaches for resource optimization,” *ACM Comput. Surveys*, vol. 51, no. 5, pp. 1–28, 2018.
[4] K. Bilal, O. Khalid, A. Erbad, and S. U. Khan, “Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers,” *Comput. Netw.*, vol. 130, pp. 94–120, Jan. 2018.
[5] S. Yi, C. Li, and Q. Li, “A survey of fog computing: Concepts, applications and issues,” in *Proc. Workshop Mobile Big Data*, 2015, pp. 37–42.
[6] K. Dolui and S. K. Datta, “Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing,” in *Proc. Global Internet Things Summit (GIoTS)*, Jun. 2017, pp. 1–6.
[7] K. Gai, M. Qiu, H. Zhao, L. Tao, and Z. Zong, “Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing,” *J. Netw. Comput. Appl.*, vol. 59, pp. 46–54, Jan. 2016.
[8] D. P. Abreu, K. Velasquez, M. Curado, and E. Monteiro, “A resilient Internet of Things architecture for smart cities,” *Ann. Telecommun.*, vol. 72, pp. 19–30, Feb. 2017.
[9] M. Jia, J. Cao, and W. Liang, “Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks,” *IEEE Trans. Cloud Comput.*, vol. 5, no. 4, pp. 725–737, Oct./Dec. 2015.
[10] K. Ha, Y. Abe, Z. Chen, W. Hu, B. Amos, P. Pillai, and M. Satyanarayanan, “Adaptive VM handoff across cloudlets,” *Dept. Comput. Sci., Carnegie Mellon Univ.*, Tech. Rep. CMU-CS-15-113, Jun. 2015.
[11] K. Wang, M. Shen, J. Cho, A. Banerjee, J. Van Der Merwe, and K. Webb, “MobiScud: A fast moving personal cloud in the mobile network,” in *Proc. 5th Workshop All Things Cellular, Oper., Appl. Challenges*, 2015, pp. 19–24.
[12] Y . Wu and L. Ying, “A cloudlet-based multi-lateral resource exchange framework for mobile users,” in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 927–935.
[13] X. Guo, L. Liu, Z. Chang, and T. Ristaniemi, “Data offloading and task allocation for cloudlet-assisted ad hoc mobile clouds,” *Wireless Netw.*, vol. 24, no. 1, pp. 79–88, Jan. 2018.
[14] M. Chen, Y. Hao, Y. Li, C.-F. Lai, and D. Wu, “On the computation offloading at ad hoc cloudlet: Architecture and service modes,” *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 18–24, Jun. 2015.
[15] R. Agarwal and A. Nayak, “DRAP: A decentralized public resourced cloudlet for ad-hoc networks,” in *Proc. IEEE 4th Int. Conf. Cloud Netw. (CloudNet)*, Oct. 2015, pp. 309–314.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com