Load Balancing Mechanism in Fog Computing to minimize workload on server

Ms. Sakshi Dinkar Junare[1], Dr. Manjiri Karande[2]
[1]Student, Department of Computer Science & Engineering, Padm. Dr. V.B. Kolte,College of Engineering, Malkapur
[1]Email: sakshijunare@gmail.com
[2]Assistant Professor, Department of Computer Science & Engineering, Padm. Dr. V. B. Kolte College of Engineering, Malkapur

DOI: 10.5281/zenodo.15751374

ABSTRACT

To address the limitations of cloud computing for latency-sensitive applications, fog computing has emerged as a contemporary distributed architecture. By extending computation and storage to the network's edge, fog systems offer location awareness, mobility support, and significantly reduced delays. Effective load balancing is crucial in fog networks to prevent uneven resource distribution among fog nodes. Optimizing load balancing can enhance Quality of Service (QoS) factors, including resource utilization, throughput, cost, response time, performance, and energy efficiency. This article provides a systematic solution of load balancing strategies in fog computing, categorizing them into approximate, exact, fundamental, and hybrid methods. Furthermore, it analyzes the metrics used for load balancing, outlining the strengths and weaknesses of each approach. The evaluation methodologies and tools employed in the reviewed studies are also examined. Finally, in this paper we identified key open challenges and future directions for load balancing mechanisms in fog computing.
Keywords: Fog computing, load balancing, Quality of Service, Metrics, Mechanisms

1. INTRODUCTION

Fog computing, a geographically distributed architecture that extends cloud capabilities, brings computational power and networking closer to end-users and IoT devices by leveraging widespread fog nodes. In traditional cloud-centric systems, data requiring processing and storage is transmitted to distant cloud servers, potentially leading to increased latency, security vulnerabilities, and reduced mobility and reliability. With the rise of location-aware and latency-sensitive applications, the inherent delays of cloud-only solutions become problematic. Fog computing addresses this by placing computational resources near IoT devices, significantly reducing latency and meeting stringent real-time requirements. By seamlessly integrating with and complementing cloud services, fog computing enables a new generation of applications.

In modern fog environments, users demand applications that are responsive and efficient. Load balancing is a critical factor in achieving high Quality of Service (QoS) in fog networks. While extensive research has been conducted on load balancing in cloud computing due to its escalating workloads, fog networks present unique challenges. Their heterogeneous and dynamic nature renders many cloud-based load balancing mechanisms unsuitable.

The objective of load balancing in fog computing is to distribute incoming workloads across available fog nodes or the cloud, utilizing a suitable strategy to prevent node overload or underload. This optimization aims to maximize throughput, performance, and resource utilization while minimizing response time, cost, and energy consumption.

2. LITERATURE SURVEY

Due to the unprecedented amount of data and the connection of over 50 billion devices to the Internet (based on Cisco estimation), handling that much of data with traditional computing models, like cloud computing, distributed computing, etc. is difficult [13]. Often privacy gaps, high communication delay, related network traffic loads that connect cloud computing to end-users for unpredictable reasons with the recent expansion of services related to IoT (like smart cities, eHealth, industrial scenarios, smart transportation systems, etc. [14]) are some challenges that affect cloud computing performance. To refer to some of cloud computing limitations and to bring cloud service traits so much closer to “Things”, as it is referred to, including cars, mobile phones, embedded systems, sensors, etc., the research community has suggested the fog computing concept [1].

Fog computing is regarded as a platform bringing cloud computing to end-users’ vicinity. “Fog”, as a term, has an analogy with real-life fog and was initially introduced by Cisco [1]. When the fog is nearer to the earth, clouds are up above in the sky and, interestingly, fog computing applies this concept, when the virtual fog platform is located closer to end users just between end-users’ devices and the cloud. In a similar definition, fog computing is suggested to make computing possible at the network edge, to send new services and applications specifically for the Internet future [15].

Bonomi, et al. [1], to give a more appropriate definition of fog computing for the first time, said that fog computing was not exclusively located at the network edge. However, it was a virtualized platform providing networking services, storage, and computations among the data centers and end devices of conventional cloud computing.

Fog computing is most often mistaken for edge computing, but we have major differences between the two. Fog computing applications are run in a multi-layer architecture that disconnects and meshes the software and hardware functions, permitting the dynamic reconfigurations for diverse applications while executing transmission services and intelligent computing. Edge computing, on the other hand, creates a direct transmission service and manages special applications in a fixed logic location. While Fog computing is hierarchical, edge computing is limited to a few peripheral devices. Besides networking and computation, fog computing deals with the control, storage, and acceleration of data-processing [16], [17]. An IoT client or smart end-device, to recognize fog computing from other computing standards, needs to utilize the following characteristics but not all of them while consuming a fog computing service [13], [18]

3. CHALLENGES IN FOG COMPUTING

Fog computing classified as the evolved extension of the cloud computing system to handle IoT related problems and shortcomings at the network edge. However, in fog computing, processing nodes are distributed and heterogeneous. Furthermore, the services based on fog technology have to work with various aspects of the restricted environment. Moreover, assurance of security is dominant in fog computing. Therefore, discovering the challenges of fog computing from service-oriented, structural, security perspectives in this technology can be listed as follows:

- **Service-Oriented**:
  Resources enrich not all fog nodes. Therefore, comprehensive scale application enhancement in resource-restricted nodes is not natural compared to traditional data centres. Therefore, distributed application development needs for potential programming platforms in Fog are required to implement. Moreover, a fog administrator is required to clarify the policies to distribute required tasks among sensors/IoT devices, fog infrastructure.
- **Structural Issues**:
  The infrastructure of fog computing consists of various components from both core and verge of networks. These types of components are equipped with a different computation but not designed for general computing. Therefore, redesign or modified the computation unit for the component is an extremely challenging part of the system setup. Additionally, Based on execution operations and operational requirements, the selection of the suitable device, places of deployment, and corresponding resource configuration are crucial in fog computing as well. Computational devices are spread across network boundaries in fog computing and can be shared or virtualised. In this case, it is necessary to define suitable metrics, strategies for inter-nodal cooperation, and efficient resource provisioning.
- **Security Aspects**:
  Fog computing rely on conventional networking components, it is highly defenceless to security attacks. Maintenance of privacy and authenticated access to computing and storage services in a widely distributed model, such as fog computing, is challenging to ensure. Therefore, Maintaining QoS is difficult during the implementation of security, where the data-centre integrity adequate and makes security topic in fog computing challenging.

4. RESEARCH OBJEVTIVES

- To understand the concept of load balancing by simulating and sending some load on server.
- To distribute the load in two different server through fog computing feature.
- To store two different files in two different server to modulate the concept of decentralized computing environment.

- To processed decentralized files (separated due to load) to provide faster solutions.

4. SYSTEM ARCHITECTURE

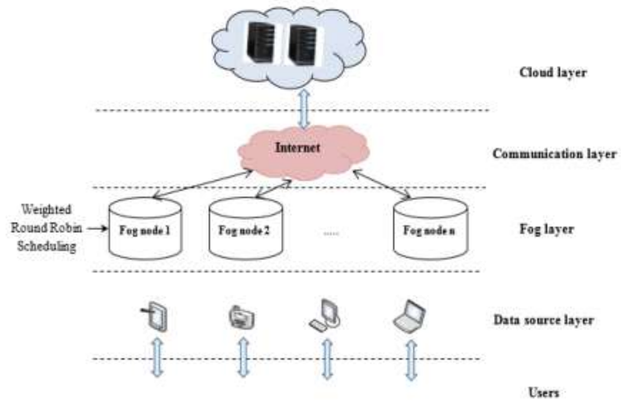

Fig.1 System Architecture

5. PROPOSED SYSTEM

Cloud load balancing is defined as dividing workload and computing properties in cloud computing. It enables enterprises to manage workload demands or application demands by distributing resources among multiple computers, networks or servers. Cloud load balancing involves managing the movement of workload traffic and demands over the Internet. Traffic on the Internet is growing rapidly, accounting for almost 100% of the current traffic annually. Therefore, the workload on the servers is increasing so rapidly, leading to overloading of the servers, mainly for the popular web servers. There are two primary solutions to overcome the problem of overloading on the server-

- First is a single-server solution in which the server is upgraded to a higher-performance server. However, the new server may also be overloaded soon, demanding another upgrade. Moreover, the upgrading process is arduous and expensive.
- The second is a multiple-server solution in which a scalable service system on a cluster of servers is built. That's why it is more cost-effective and more scalable to build a server cluster system for network services.
- Cloud-based servers can achieve more precise scalability and availability by using farm server load balancing. Load balancing is beneficial with almost any type of service, such as HTTP, SMTP, DNS, FTP, and POP/IMAP.

6. FLOW OF THE SYSTEM

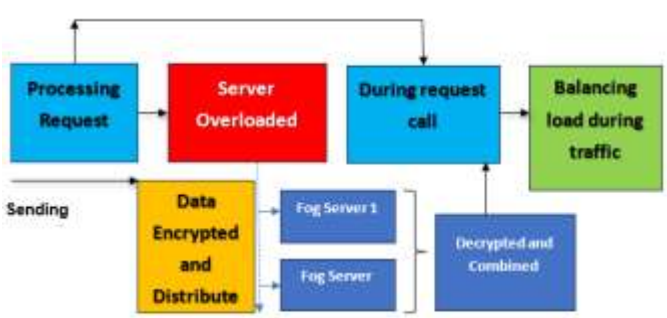

Fig. 2 Flow of the system

7. RESULT

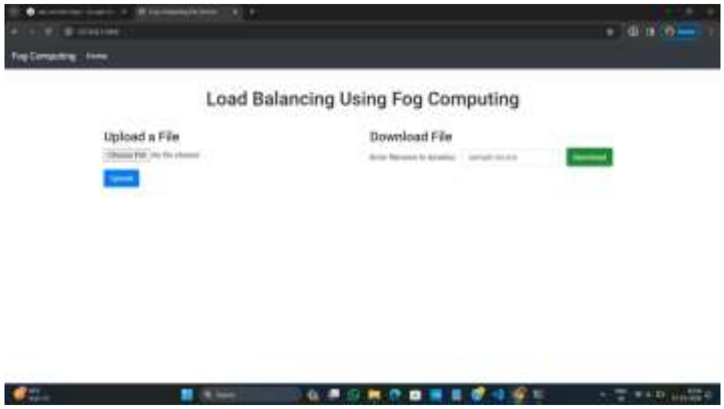

Fig. 3 Dashboard for uploading files on server

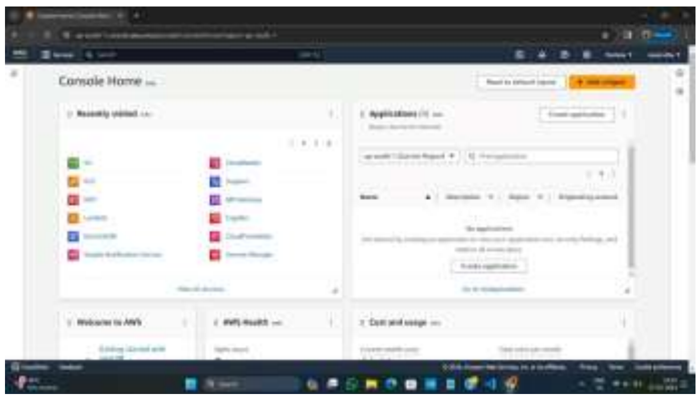

Fig. 4 Two server created for sharing the data

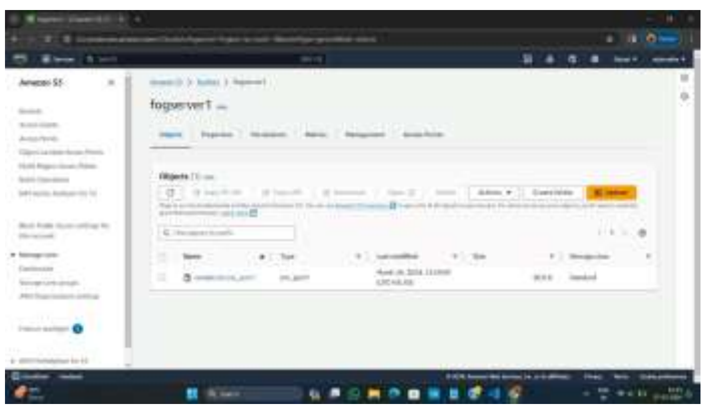

Fig. 5 After uploading file encrypted and shared

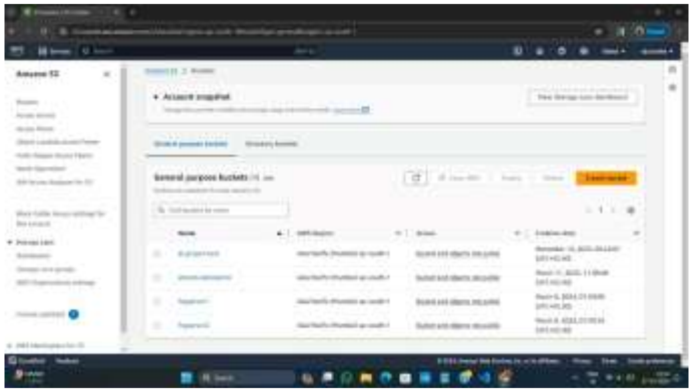

Fig. 6 During decryption file was restored

8. CONCLUSION

The distribution of services and computational resources across cloud and edge networks continues to be a central focus of research in both industry and academia due to its significant potential. This work addresses the task scheduling challenges within this evolving environment, aiming to ensure efficient task execution based on available processing power and energy resources. Modern web applications frequently experience concurrent access from numerous users, creating substantial load management difficulties that can lead to system instability. To mitigate this issue, our proposed system leverages fog computing to dynamically balance workloads, thereby facilitating smoother request processing and reducing latency.

REFERENCES

[1] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in Proceedings of the first edition of the MCC workshop on Mobile cloud computing, 2012: ACM, pp. 13-16.

[2] A. Yousefpour et al., "All one needs to know about fog computing and related edge computing paradigms: A complete survey," Journal of Systems Architecture, vol. 98, pp. 289-330, 2019/09/01/ 2019.

[3] N. Auluck, A. Azim, and K. Fizza, "Improving the Schedulability of Real-Time Tasks using Fog Computing," IEEE Transactions on Services Computing, pp. 1-1, 2019.

[4] S. Aslam and M. A. Shah, "Load balancing algorithms in cloud computing: A survey of modern techniques," in 2015 National Software Engineering Conference (NSEC), 2015: IEEE, pp. 30-35.

[5] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," Journal of Systems and Software, vol. 80, no. 4, pp. 571-583, 2007/04/01/ 2007.

[6] P. Jamshidi, A. Ahmad, and C. Pahl, "Cloud Migration Research: A Systematic Review," IEEE Transactions on Cloud Computing, vol. 1, no. 2, pp. 142-157, 2013.

[7] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," Information and Software Technology, vol. 51, no. 1, pp. 7-15, 2009/01/01/ 2009.

[8] C. Jatoth, G. R. Gangadharan, and R. Buyya, "Computational Intelligence Based QoS-Aware Web Service Composition: A Systematic Literature Review," IEEE Transactions on Services Computing, vol. 10, no. 3, pp. 475-492, 2017.

[9] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, "Requirements engineering paper classification and evaluation criteria: a proposal and a discussion," Requirements engineering, vol. 11, no. 1, pp. 102–107, 2005.

[10] M. Haghi Kashani, A. M. Rahmani, and N. Jafari Navimipour, "Quality of service-aware approaches in fog computing," International Journal of Communication Systems, 2020/01/10 2020.

[11] M. Rahimi, M. Songhorabadi, and M. H. Kashani, "Fog-based smart homes: A systematic review," Journal of Network and Computer Applications, vol. 153, p. 102531, 2020/03/01/ 2020.

[12] S. Bazzaz Abkenar, M. Haghi Kashani, E. Mahdipour, and S. M. Jameii, "Big data analytics meets social media: A systematic review of techniques, open issues, and future directions," Telematics and Informatics, p. 101517, 2020/10/14/ 2020.

[13] P. Hu, S. Dhelim, H. Ning, and T. Qiu, "Survey on fog computing: architecture, key technologies, applications and open issues," Journal of network and computer applications, vol. 98, pp. 27-42, 2017.
[14] P. Asghari, A. M. Rahmani, and H. H. S. Javadi, "Internet of Things applications: A systematic review," Computer Networks vol. 148, pp. 241-261, 2019/01/15/ 2019.
[15] E. Marín-Tordera, X. Masip-Bruin, J. García-Almiñana, A. Jukan, G.-J. Ren, and J. Zhu, "Do we all really know what a fog node is? Current trends towards an open definition," Computer Communications, vol. 109, pp. 117-130, 2017.
[16] O. C. A. W. Group, "OpenFog reference architecture for fog computing," OPFRA001, vol. 20817, p. 162, 2017.
[17] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, "A comprehensive survey on fog computing: State-of-the-art and research challenges," IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 416-464, 2017.
[18] M. Iorga, L. Feldman, R. Barton, M. Martin, N. Goren, and C. Mahmoudi, "Fog computing conceptual model, recommendations of the National Institute of Standards and Technology," NIST Special Publication, pp. 500-325, 2018.