JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A SYSTEMATIC REVIEW ON LOAD BALANCING MECHANISM IN FOG COMPUTING

[1]Miss. Sakshi Dinkar Junare, [2]Prof. M. U. Karande
[1]Student, [2]Assit. Professer,
[1]Computer Engineering,
[1]Padmashri Dr. V.B. Kolte College of Engineering, Malkapur, India

Abstract : Recently, fog computing has been introduced as a modern distributed paradigm and complement to cloud computing to provide services. Fog system extends storing and computing to the edge of the network, which can solve the problem about service computing of the delay-sensitive applications remarkably besides enabling the location awareness and mobility support. Load balancing is an important aspect of fog networks that avoids a situation with some under-loaded or overloaded fog nodes. Quality of Service (QoS) parameters such as resource utilization, throughput, cost, response time, performance, and energy consumption can be improved with load balancing. In recent years, some researches in load balancing techniques in fog networks have been carried out, but there is no systematic review to consolidate these studies. This article reviews the load balancing mechanisms systematically in fog computing in four classifications, including approximate, exact, fundamental, and hybrid methods. Also, this article investigates load balancing metrics with all advantages and disadvantages related to chosen load balancing mechanisms in fog networks. The evaluation techniques and tools applied for each reviewed study are.

Keywords: Fog computing, load balancing, Quality of Service, Internet of things, systematic review

I. INTRODUCTION

Fog computing, which extends from the cloud and is a geographically distributed paradigm, brings networking power and computing into the network edge, closer to end-users and IoT devices both because of being supported by wide-spread fog nodes. Most of the data, in cloud-only architectures, requiring processing, analysis, and storage, are transmitted to the cloud servers, which may have an influence on latency, security, mobility, and reliability adversely. With the existence of location-aware and delay-sensitive applications, the cloud on its own comes across some problems to meet the extremely-low latency requirements of these applications; the proximity of the fog layer to the Internet of Things (IoT) devices may remarkably decrease latency and meet the needs of extremely-low . Fog computing always interacts with and supports the cloud, creating a novel generation of applications and services. Nowadays, in fog computing environments, users need applications that give quick responses whenever they want to access anything and work fast. To improve QoS factors in a fog network significantly, we can use an efficient load balancing strategy because load balancing is regarded as an important issue. Many studies have been done to balance the cloud computing load because the load on the cloud increases enormously [4]. Being heterogeneous and dynamic, the fog networks cannot directly apply most of the load balancing mechanisms of cloud computing; The goal of load balancing in fog environment is to distribute the coming load between available fog nodes or cloud, based on one mechanism, to avoid overload or under-load of fog nodes. This mechanism can 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 . 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.) 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 .

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 . 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 .

Bonomi, et al, 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. 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 .

3. Fog Architecture

Fog computing architecture reference model is an important study topic. Recently, a wide range of architectures has been suggested for fog computing, mostly obtained from a structure with three layers. Fog network expands cloud services to the network edge by suggesting a fog layer between cloud and user devices. As it can be observed, Fig. 1 illustrates the fog architecture hierarchically, having three layers as follows:

- Cloud layer: The layer of cloud computing is composed of different storage devices and high-performing servers and creates several services of applications. It bears robust storage and computing abilities to back the permanent storage of a large amount of information and extended computation analysis. However, it should be noted that all computing and storage tasks do not pass the cloud that is not the same as traditional cloud computing architecture.
- Fog layer: The fog layer is located at the network edge, which consists of a couple of fog nodes like access points, routers, switches, gateways, etc. They are spread between cloud and end devices. In order to get services, end devices might easily connect with fog nodes. They are capable of computing, storing, and transmitting the received sensed data. The latencysensitive applications and real-time analysis can be performed in a fog layer. In addition, we can refer to the connection between the cloud data center and the fog nodes by the IP core network. In order to get more robust storage and computing capabilities, fog nodes have the responsibility of interacting and cooperating with the cloud.
- User device: The layer of a user device is so close to the physical environment and end-user. This layer is composed of different IoT devices, like, sensors, cellphones, smart automobiles, cards, and readers. Although cellphones and smart vehicles have got computing capabilities, they are utilized as just smart sensing devices. Generally and geographically, they are widely distributed and

responsible for sensing feature data related to events or physical objects and for transferring them to upper layers to be processed and stored. Here in the architecture, all end devices or smart objects are connected with fog nodes by technologies with wired or wireless connection access such as 3G, 4G, wireless LAN, ZigBee, Bluetooth, and Wi-Fi. Wireless or wired communication technologies to help the interconnection and intercommunications of fog nodes. IP core network helps fog nodes each to be linked with the cloud .

4. Load Balancing

In fog system, load balancing facilitates the distribution of workload on resources equally, aiming to provide services continually if the service component fails, and it is done by provisioning and de-provisioning instances of applications along with proper resource utilization. Because data centers procure diversities between hosts and show special features of traffic, an appropriate mechanism of load balancing is needed in fog computing to refine the performance of applications and utilization of the network. To evade any overload or under-load on resources, load balancing, as a mechanism, spread the workload onto different resources. Load balancing, which distributes the load among different resources, is implemented either in physical equipment or software. The load balancing has some goals, including throughput maximization, response time minimization, and traffic optimization. Consumption optimization in the server-side resources, request processing time minimization, and scalability improvement in the distributed environment are some other purposes of the technique of load balancing. In fog networks, load balancing may have various methods that can be of static or dynamic nature or both. In static methods, with primary information about the system as a necessary feature, the rule should be programmed in the load balancer because the user's behavior is not predictable, and methods of static load balancing are not necessarily efficient in the network. Further, the dynamic methods outperform the static methods because of the dynamic distribution of load based on the pattern that is programmed in the load balancer.

Mechanisms of dynamic load-balancing apply current system state to this end, and they use especial policies including:

- Transfer: It defines the conditions based on which a task has to be sent from one node to the other. The arriving tasks that enter the transfer policy are transferred or processed based on a determining rule that relies on each node workload. The policy deals with task migration and rescheduling.
- Selection: It defines whether a task should be sent or not and also regards a couple of elements to select a task, like the amount of overhead needed for migration, time of task execution, and total nonlocal system calls.
- Location: It defines under-loaded nodes and then sends tasks to these nodes. In aimed nodes, the availability of essential services for task rescheduling or migration is checked.
- Information: The complete information, considering system nodes, is collected in this policy and is used by other policies to make a decision. This policy determines the time at which the information should be collected. Various policies have some relationships that are mentioned below: Transfer policy grabs incoming tasks and determines whether to transfer them to a remote node to balance load or not. If not eligible to be transferred, the task will be locally processed. When the transfer policy decides to transfer a task, the location policy would be triggered to locate a remote node to process the task. The task is locally processed when a far partner cannot be found, or the task is transferred to a remote node.

5. 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.

6. System Architecture

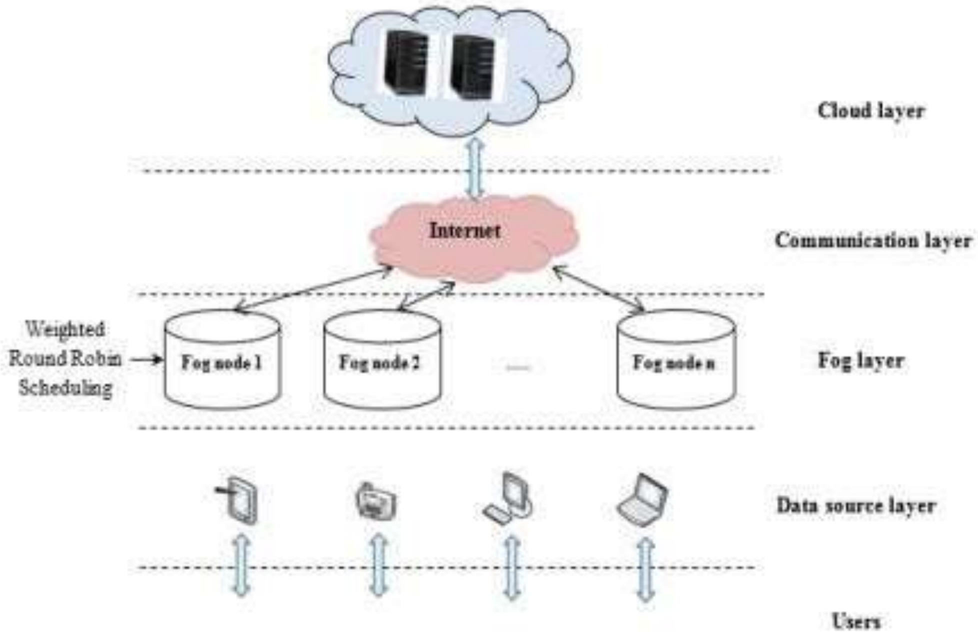

Fig.1 System Architecture

7. 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.

IV. Conclusion

In the 21st century, the FC paradigm is expected to remain of interest for the researchers in industry and academia provide its incredible potential where services and computing resources are distributed in effective FNs reside at the cloud computing network edge. In this paper, we concentrated on the task scheduling problem in the environment of FC to assure the effective task execution according to the available processing capacity and remaining energy. Thousands of people have access to a website at a particular time. This makes it challenging for the application to manage the load coming from these requests at the same time. Sometimes this can lead to system failure.

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] 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