LOAD BALANCING METHODS IN CLOUD INFOCOMMUNICATION SYSTEMS

Samad Habib Suhel, Saburova S.A.,Kadatskaja O.I.

Кафедра «Інфокомунікаційної інженерії»
ім. В.В. Поповського», ХНУРЕ, Україна

E-mail: svitlana.saburova@nure.ua

Abstract

Balancing methods considered on base Model of Load Balancing Metrics, LB algorithms that broadly classified as static, dynamic, and hybrid on information about the system state of the infrastructure on which the algorithms based. Presented Model of Load Balancing Algorithms on the basis of nature and state of the system infrastructure, Model of load balancing algorithms based on the attribute used, Model of load balancing on the basis functionality and type, Model of load balancing algorithms on the basis of infrastructure with the MapReduce algorithm applied to simulation results and simulation of dynamic load balancing using traffic route optimization on base Traffic Route Optimization.

With the active development of high-speed Internet technologies, data and program storage provided through the access infrastructure on cloud computing platforms instead of the hard disk of a personal computer and servers of corporate and local networks. In cloud computing, the elements involved are clients, data centers and distributed servers.

To meet the requirements of the quality management system, QoS (Quality of Service) is load balancing in the infrastructure and cloud-computing platform. Load balancing allows for even distribution of workload during business time across multiple sites so that no site overloaded with explosive traffic.

Existing research and development of load balancing architecture and algorithms allow solving the problems of traffic transmission to cloud computing platforms within the normalized quality parameters by minimizing the response time within the requirements of service classes in the QoS system.

The idea of creating cloud platforms formed during the advent of the Internet, but the first cloud platform appeared only in 2006 under the name
Amazon Web Services (although the company began providing access to computing resources via the Internet in 2002). In 2006, Amazon offered more than 50 different services in 14 geographic regions.

The next big platform was the Microsoft Azure system, which appeared in 2010 (now ranks second in the number of users and services provided, after Amazon).

In 2011, the world's third largest player in the cloud computing market was introduced - Google Cloud Platform. A year later, the active creation of new platforms and services from various companies began, and because of lower platform rental prices. Services such as i-Teco OpenStack Cloud, Azure Cloud OS, Jelastic for PHP and Java applications and others created.

Today, cloud platforms are more popular than ever. The main advantages of using cloud platforms are the speed of creating new applications, flexibility and scalability of the system. The top 10 most famous and used cloud platforms include: Amazon Web Services, Azure from Microsoft, Google App Engine, Rackspace, Force.com from Salesforce, Intuit Partner Platform, Facebook, IBM Cloud, VMWare vCloud, Sharepoint Online. In 2020 year – near 94% of Internet users used free and unlimited cloud storage.

Cloud computing conceptual reference model identifies the major actors, their activities and functions in cloud computing.

Figure 1 presents an overview of the cloud reference architecture.

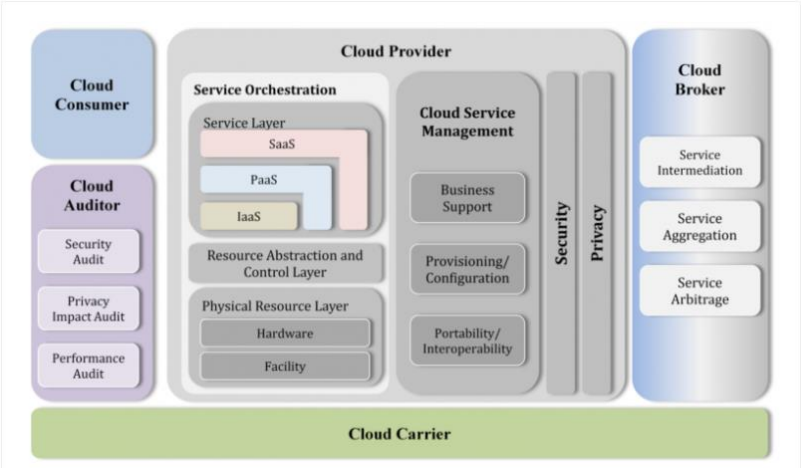

Fig.1.The conceptual reference model

As shown in fig. 1, the cloud computing reference architecture defines five major actors: cloud consumer, cloud provider, cloud carrier, cloud auditor and cloud broker. Each actor is an entity (a person or an organization) that participates in a transaction or process and/or performs tasks in cloud computing.

On Table 1 shows actors in cloud computing.

Table 1 - Actors in cloud computing

SNO	Actor	Definition
1	Cloud Consumer	A person or organization that maintains a business relationship with, and uses service from, Cloud Providers
2	Cloud Provider	A person, organization, or entity responsible for making a service available to interested parties.
3	Cloud Auditor	A party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation.
4	Cloud Broker	An entity that manages the use, performance and delivery of cloud services, and negotiates relationships between Cloud Providers and Cloud Consumers.
5	Cloud Carrier	An intermediary that provides connectivity and transport of cloud services from Cloud Providers to Cloud Consumers.

Figure 2 illustrates the interactions among the actors in cloud computing. A cloud consumer may request cloud services from a cloud provider directly or via a cloud broker. A cloud auditor conducts independent audits and may contact the others to collect necessary information.

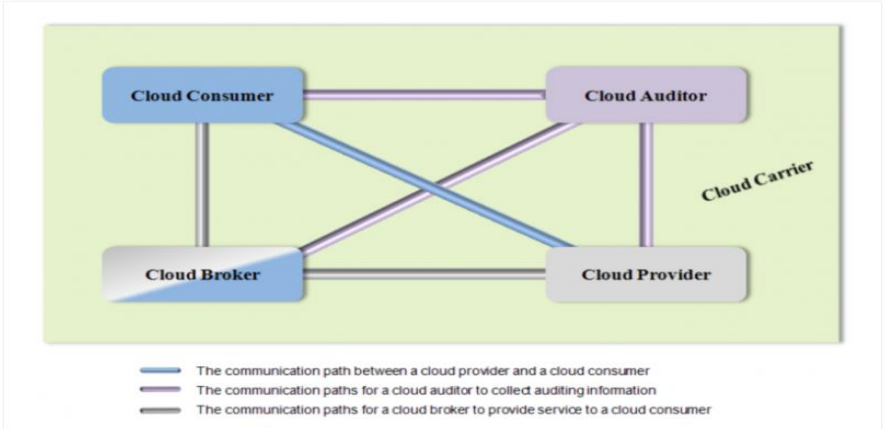

Fig.2. Interactions among the actors in cloud computing

An undesirable event on the CSP (Configuration Service Provider) side that reduces the performance and efficiency of computing resources along with guaranteed quality of service (QoS) in accordance with an agreed service level agreement (SLA) between the customer and the supplier is the problem of unbalanced load during the busy time of the infrastructure as a client and the supplier.

At the physical machine or virtual machine level, load balancing in cloud computing can be performed during peak hours to avoid explosive traffic [1].

When a group of tasks arrives at the virtual machine and the resources in the BT are maximally loaded, this causes requests for additional resources.

The VM enters an overload state. Then the tasks will either be blocked or become deadlocked without success on their execution. Therefore, it becomes necessary to transfer tasks to another resource and another virtual machine.

This workload migration process involves three main steps:

- load balancing, which controls the current load on the machine resource,
- provides additional resource discovery that finds another suitable resource;
- workload migration, which moves additional tasks to available resources.

These processes handled by three different load balancers, resource discovery, and task migration, respectively, during busy hours.

Therefore, no computing machine overloaded, under loaded or idle in the process of optimal redistribution of the workload in a distributed system, such as cloud computing, under conditions of guaranteed load balancing in BT [13].

Load balancing will speed up various rated parameters;

– response time,
– lead time,
– system stability, etc. [2].

This method of scheduling tasks is a complex optimization task.

In planning and distributing traffic, scheduling, allocating and managing resources, researchers have proposed many approaches to methods of load balancing in BT. In this study, the main factors responsible for the problem of load imbalance indicated, and methods for their reduction in BT are proposed. Processes of two-tier model in unbalanced clouds.

A two-tier model in unbalanced clouds proposed because of a modified architecture to provide a standardized load reduction in BT, as shown in fig. 3 [2].

Note that the Virtual Machine Manager and Virtual Machine Monitor abstracted in this model.

At the physical machine (PM) level, the first level load balancing performed, and the second level performed at the virtual machine level in the BT.

As a result, two kinds of migration tasks implemented: inside VM – task migration; between VMs – task migration.

Process of inside VM – task migration. The query complex generates custom queries, which are custom tasks that require computational resources to complete. The data center controller is responsible for managing tasks. The choice of a virtual machine for a given user task is decided by the BT load balancer. On separate physical machines, distributing traffic between the corresponding connected virtual machines, the first-level load balancer balances this traffic in the BT.

Processes between VMs – task migration. A tier 2 load balancer distributes production traffic between different virtual machines on different physical machines.

Scheduling and distributing tasks among virtual machines based on QoS system requirements constitutes the cloud computing workload. The load balancing process includes some steps.

In figure 3 shows, two-tier model with two levels of balances traffic in the BT.

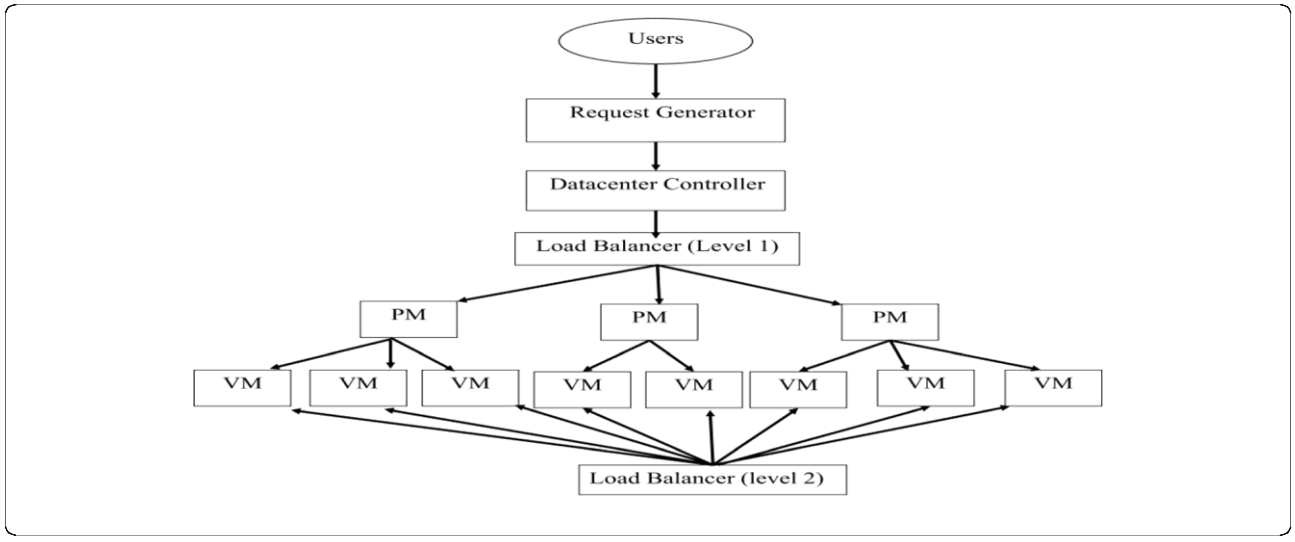

Fig.3. Two-tier model with two levels of balances traffic in the BT

Composition of requirements for a custom task. At this point, resources for custom tasks scheduled for the virtual machine to run. Process for identifying virtual machine resource information. At this stage, the state of the virtual machine can be defined as balanced, overloaded or under loaded relative to the threshold value in BT. The status of the virtual machine resource information is checked. Controls the status of the current used of virtual machine resources and unallocated resources. Planning specific tasks of providing traffic resources in BT. Once the resource information for each virtual machine established, tasks are scheduled using a scheduling algorithm to load the resources on the corresponding virtual machines in the WT. Distribution of traffic resources in BT. Resources for traffic in BT allocated for the execution of specifically scheduled tasks. In this case, the policy of allocating normalized resources is used. A large number of scheduling and allocation policies proposed in the literature. As a rule, planning is required to accelerate the implementation of the provision of traffic with resources, the distribution policy used to effectively and efficiently manage resources and increase their productivity in BT. The efficiency of the load-balancing algorithm determined by the metric of the levels of efficiency of the scheduling algorithm and the distribution policy [2]. Process of migration. In the process of load balancing in the cloud, migration is an important step. As noted earlier, there are two types of migrations in the cloud, depending on the types: virtual machine migration and task migration.

The results shown in fig. 4.

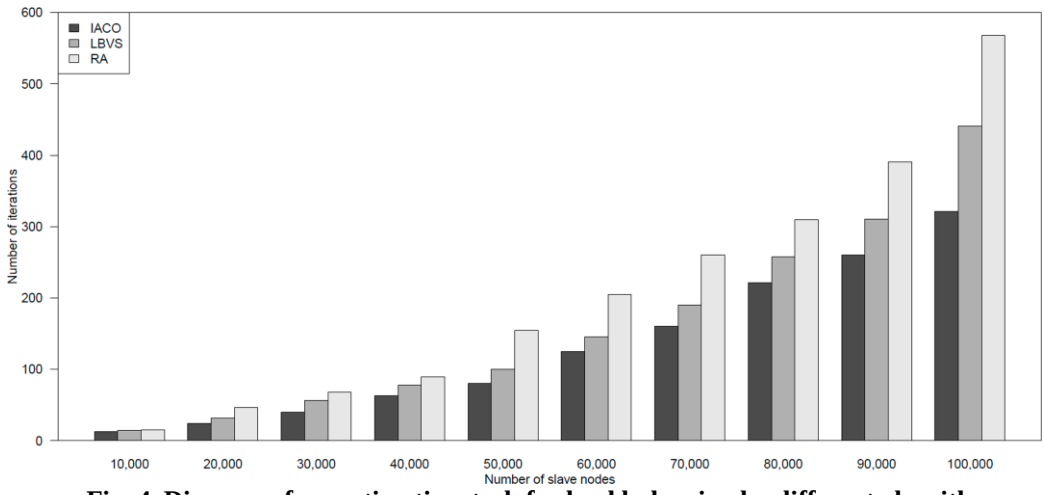

Fig. 4. Diagram of execution time task for load balancing by different algorithms

Conclusion

The process of moving a virtual machine from one physical host to another to get rid of the overload problem, in real time and migrations without using virtual machines is a virtual machine migration. The process of moving tasks between virtual machines is of two types: migrating tasks within virtual machines and migrating tasks between virtual machines is a task migration. Thus, the task migration process is more time and cost efficient than virtual machine migration, and the trend has shifted from virtual machines to task migration [3].

To balance the load with the number of slave nodes from 10.000 to 100.000, calculations, analysis and assessment of the possibility of working with different scales of slave nodes in the cloud-computing platform carried out.

Based on the proposed solutions to load balancing cloud computing and resource infrastructure, it should be borne in mind that so many service providers who are responsible for supporting an application in the cloud. Google, Amazon, Microsoft and many more are the most common cloud service providers. These service providers maintain data centers around the world where data is stored in large quantities and requests processed to serve the huge global traffic. Classification of task scheduling and load balancing algorithms.

The classification of task scheduling and load balancing algorithms in BT includes seven different categories. Performed simulation of dynamic load balancing with use of traffic route optimization. In cloud computing, load balancing plays an important role in providing quality of service (QoS) guarantees. Therefore, a new approach to load balancing proposed using Business Time Traffic Route Optimization (TMO) dynamically balance the business time workload on a cloud-computing platform.

References:

1. Mahmood Z (2011, August) Cloud computing: characteristics and deployment approaches. In: 2011 IEEE 11th international conference on Computer and Information Technology (CIT), pp 121–126.

2. Jain N, Choudhary S (2016, March) Overview of virtualization in cloud computing. In: Symposium on colossal data analysis and networking (CDAN), pp 1–4.

3. Standard ISO 17788 2014 - Information technology - Cloud computing - Overview and vocabulary, pp 1-10.