A Survey on Routing Algorithms for Cloud Computing

K.Kuppusamy, Ph.D
Professor
Department of Comp.Sci & Engg
Alagappa University – Karaikudi.

J.Mahalakshmi
Lecturer
Department of IT
Dr. Umayal Ramanathan College For Women - Karaikudi

ABSTRACT

Cloud computing is an evolving paradigm, as the NIST defines cloud computing is the internet based computing where the shared servers virtually provides service, software, infrastructure, platform, and other devices and resources and hosting to the customers on a Pay – as – you – use basis. It also enables the convenient, on – demand network access to a shared pool of configuration enable computing resources. Cloud computing appears as a computational paradigm as well as a distributed architecture, and its main objective is to provide secure as well as quick computing service to the consumers. The cloud enhances the agility, scalability, ability to adapt the adequate number of users according to the demand cost reduction using the optimized and efficient computing. In this paper, we surveyed the various routing algorithms that are used for the cloud computing processes, the optimal resource allocation techniques used in cloud computing and its applications in various fields. Cloud computing is widely used in distributed and mobile computing environment. This is possible due to miniaturization of communication technology. The significance of the routing is considered as an important part in the cloud computing since they are based on the on – demand networks. Hence allocating the nearest route is a vital role in cloud computing.

General Terms

Routing Algorithms, Optimization Algorithms.

Keywords

Cloud Computing, Routing Algorithms, Distributed System, Resource Allocation Techniques, Optimization.

1. INTRODUCTION

Cloud computing can be defined by the NIST as a model for enabling convenient, on – demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [9]. Cloud computing can be considered as a new computing paradigm with implications for greater flexibility and availability at lower cost. The term itself is often used today with a range of meanings and interpretations [9]. The cloud model is composed of five essential characteristics, three service models and four types of deployment models.

Cloud computing appears as a computational paradigm as well as a distribution architecture and its main objective is to provide secure, quick, convenient data storage and computing services, with all computing resources visualized as services and delivered over the internet [10]. The cloud enhances the collaboration, agility, scalability, availability, ability to adapt to fluctuations according to the demand, accelerate development works and provides potential for cost reduction through optimized and efficient computing.

Recently, a number of commercial and academic organizations have built large systems from commodity computers, disks, and networks, and have created software to make this hardware easier to program and manage. These organizations have taken a variety of novel approaches to address the challenges outlined above. In some cases these organizations have used their hardware and software to provide storage, computational, and data management services to their own internal users, or to provide these services to external customers for a fee.

Normally it refers to the hardware and software environment that implements this service-based environment as a cloud-computing environment. Because the term "cloud computing" is relatively new, there is not universal agreement on this definition. Some people use the terms grid computing, utility computing, or application service providers to describe the same storage, computation, and data-management ideas that constitute cloud computing.

Cloud computing combines a number of computing concepts and technologies such as Service Oriented Architecture (SOA), Web 2.0, Virtualization, and other technologies, while their software and data are stored on the servers .Cloud computing architecture consists of two components the front end and the back end. The front end of the cloud computing comprises the client's devices or may be a computer network. Back end refers to the cloud itself which may encompass various computer machines, data storage system and servers.

Group of these cloud make a whole cloud computing system. The whole system is administered via a central server that is also used for monitoring clients demand and traffic ensuring smooth functioning of the system. A special type of software called as the "Middleware" is used to allow the computers that are connected on the network to communicate with each other. Cloud computing systems also must have a copy of all its clients' data to restore the service which may arise due to a device breakdown. Making copy of a data is called redundancy and cloud computing providers provide the data redundancy.

The rest of the paper is organized as follows. Section 2 presents the overview of the cloud computing, their deployment models, standards and essential characteristics. Section 3 defines the various routing algorithms presented for the computing.. In Section 4 some conclusion is discussed based on the routing algorithms for the cloud computing.

2. CLOUD COMPUTING OVERVIEW

The primary business service models employed are the Software as a Service. Platform as a Service and Infrastructure as a service, and common deployment models employed by service providers and users to use and maintain the cloud services such as the private cloud, public cloud , community cloud and the hybrid cloud

2.1 Essential Characteristics

2.1.1 On – demand self service - A consumer can unilaterally have the provision for computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

2.1.2 Broad network access - Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms.

2.1.3 Resource pooling - The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction. Examples of resources include storage, processing, memory, and network bandwidth

2.1.4 Rapid elasticity - Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

2.1.5 Measured service - Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized.

2.2 Service Models

2.2.1 Software as a Service (SaaS) - The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

2.2.2 Platform as a Service (PaaS) - The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

2.2.3 Infrastructure as a Service (IaaS). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications and possibly limited control of select networking components (e.g., host firewalls).

2.3 Deployment Models

2.3.1 Private cloud - The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers. It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

2.3.2 Community cloud - The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns. It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

2.3.3 Public cloud - The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

2.3.4 Hybrid cloud - The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

2.4 Benefits

The following are some of the possible benefits for those who offer cloud computing-based services and applications:

2.4.1 Cost Savings — Companies can reduce their capital expenditures and use operational expenditures for increasing their computing capabilities. This is a lower barrier to entry and also requires fewer in-house IT resources to provide system support.

2.4.2 Scalability/Flexibility - Companies can start with a small deployment and grow to a large deployment fairly rapidly, and then scale back if necessary. Also, the flexibility of cloud computing allows companies to use extra resources at peak times, enabling them to satisfy consumer demands.

2.4.3 Reliability - Services using multiple redundant sites can support business continuity and disaster recovery.

2.4.4 Maintenance - Cloud service providers do the system maintenance, and access is through APIs that do not require application installations onto PCs, thus further reducing maintenance requirements.

2.4.5 Mobile Accessible - Mobile workers have increased productivity due to systems accessible in an infrastructure available from anywhere.

3. ROUTING ALGORITHMS OVERVIEW

It is a new era of referring to access shared computing resources. In this section the survey of various routing algorithms that are employed in the wireless sensor networks, cloud communication link, optical network for the grid and the cloud computing applications, unicast and multicast routing algorithms used for the purpose, the shadow – routing algorithm based on dynamic traversal, and other algorithms and their significance is discussed.

3.1 Sensor Networks for Cloud Computing

A wireless sensor network (WSN) consists of spatially distributed autonomous sensors. Each node in a sensor network is typically equipped with a radio transceiver or other wireless communications device, a small microcontroller, and an energy source, usually a battery. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and bandwidth [3].

A sensor network is a computer network Composed of a large number of sensor nodes. [3] The development of sensor networks requires technologies from three different research areas: sensing, communication, and computing (including hardware, software, and algorithms). Routing protocols in WSNs are broadly divided into two categories: Network Structure based and Protocol Operation based. Network Structure based routing protocols are again divided into flat-based routing, hierarchical-based routing, and location-based routing. Protocol Operation based are again divided into Multipath based, Query based, QoS based, Coherent based and Negotiation based. In multipath routing, communication among nodes uses multiple paths to enhance the network performance instead of single path. In Query based routing, the destination nodes propagate a query for data from a node through the network and a node having this data sends the data which matches the query back to the node, which initiates the query.

In location-based routing, sensor nodes positions are exploited to route data in the network. In this kind of routing, sensor nodes are addressed by means of their locations. The distance between neighboring nodes can be estimated on the basis of incoming signal strengths. Relative coordinates of neighboring nodes can be obtained by exchanging such information between neighbors [17,18].Combining WSNs with cloud makes it easy to share and analyze real time sensor data on-the-fly. It also gives an advantage of providing sensor data or sensor event as a service over the internet [14]. The terms Sensing as a Service (SaaS) and Sensor Event as a Service (SEaaS) are coined to describe the process of making the sensor data and event of interests available to the clients respectively over the cloud infrastructure.

Merging of two technologies makes sense for large number of application. Some of the applications are weather forecasting, transport monitoring, healthcare etc.,

3.2 Optimal Resource Allocation Algorithms

A stochastic model of cloud computing, where jobs arrive according to a stochastic process and request resources like CPU, memory and storage space is considered in using the cloud clusters[5].It also consider a model where the resource allocation problem can be separated into a routing or load balancing problem and a scheduling problem. Also the join-the-shortest-queue routing and power of- two-choices routing algorithms with MaxWeight scheduling algorithm are also discussed. The result of the algorithms shows the efficient throughput and makes it optimal.

The paper [5] focus on cloud computing platforms that provide infrastructure as service. Users submit requests for resources in the form of virtual machines (VMs). Each request specifies the amount of resources it needs in terms of processor power, memory, storage space, etc.. The user calls these requests as jobs. The cloud service provider first queues these requests and then schedules them on physical machines called servers.

The simplest architecture for serving the jobs is to queue them at a central location. In each time slot, a central scheduler chooses the configuration at each server and allocates jobs to the servers, in a preemptive manner. As pointed out in [15], this problem is then identical to scheduling in an adhoc wireless network with interference constraints. It was shown in [10] that join-the-shortest queue-type algorithms for routing, along with the MaxWeight scheduling algorithm at each server is throughput optimal.

In this work the authors considered a stochastic model for load balancing and scheduling in cloud computing clusters. They studied the performance of Join the Shortest Queue routing and MaxWeight scheduling policy. It was known that this policy is throughput optimal. Also shown that it is heavy traffic optimal when all the servers are identical and found that using the power-of-two-choices routing instead of JSQ routing is also heavy traffic optimal.

3.3 Effective Cost Management System through MSBE

In Cloud Computing Architecture, Brokers are responsible to provide services to the end users. An Effective Cost Management System (ECMS) which works over Secure Cloud Communication Paradigm (SCCP) helps in finding a communication link with overall minimum cost of links.

Two algorithms are included, first is Secure Optimized Route Cost Finder (S-ORCF) to find optimum route between broker and cloud on the behalf of cost factor and second is Secure Optimized Route Management (S-ORM) to maintain optimum route. These algorithms proposed with cryptographic integrity of the secure route discovery process in efficient routing approaches between broker and cloud.

There is lack in Dynamic Source Routing Approach to verify whether any intermediate node has been deleted, inserted or modified with no valid authentication. The work [8] use symmetric cryptographic primitives, which is made possible due to multisource broadcast encryption scheme. This algorithm outlines the use of secure route discovery protocol (SRDP) that employs such a security paradigm in cloud computing. It is attained by directing every node to retain a reliable One-hop Reliable Delivery Neighborhood(RDN)and providing a secret to each node residing in the RDN and broadcast encrypted message that is not accessible by any other node that are not in the RDN.

3.4 Optical Clouds

An evolution towards "optical clouds", stressing the important role that especially optical networking technology can play in realizing next generation cloud solutions. Routing and path computation algorithms are undeniably of fundamental concern in communication networks. The anycast and multicast routing algorithms in general and in optical networks in particular is considered here. It will become clear how anycast routing helps to meet the requirements for scalability (load balancing) and cloud exploits geographical awareness (e.g. sending to closest candidate destination of anycast-set).

Traditionally, data networks employ unicast routing algorithms (e.g. shortest path routing) for transferring data from source to a given destination. In cloud and grid networks however, each user-generated task can be serviced at multiple locations in the network. Moreover, the exact service location and network route is of less importance to the end user.A fundamental concept to realize such service-oriented networks are the anycast routing principle. Anycast routing specifically enables users to transmit data for processing, storage or service delivery, without assigning an explicit destination. By simply using an anycast address, service providers can offer a generic interface to end users for a wide range of services and applications.

Next to anycast, the work also considers multicast routing and its application in grid or cloud networks. It can obviously be very useful to, for instance, distribute an identical data set to multiple computing nodes. Alternatively, computing and streaming of data to multiple end users of concern in case of, for example, video transcoding and scene rendering in gaming environments. Optical networks in particular may benefit from the use of anycast or multicast routing algorithms.

4. CONCLUSION

In this proposed survey work, we discusses about the basics of the cloud computing and various routing algorithms that have been employed yet for the cloud networks to improve their performance. In these algorithms the JSQ algorithm shows the optimal throughput and the shortest path scheduling algorithm shows the neighborhood route to reach. The optical networks use the multicast routing whereas the wireless sensor networks uses the optimization techniques.

As a consequence, this works can be further enhanced by implementing the new algorithm for the optimized routing service in the deployment models itself. Hybrid models can be proposed such that they can be able to increase the throughput without delay. Also the works above focuses on the software services the future implementation can be made within the infrastructure itself.

5. REFERENCES

[1] C. Ulmer, L. Alkalai and S. Yalamanchili, Wireless distributed sensor networks.

[2] Römer, Kay; Friedemann Mattern (December 2004), "The Design Space of Wireless Sensor Networks", *IEEE Wireless Communications* 11 (6): 54–61, doi:10.1109/MWC.2004.1368897,

[3]Akyildiz, I.F., W. Su, Y. Sankarasubramaniam, E. Cayirci, "A Survey on Sensor Networks", IEEE Communications Magazine, August, 102-114(2002).

[4] M. Chu, H. Haussecker, and F. Zhao, Scalable Information-Driven Sensor Querying and Routing for ad hoc Heterogeneous Sensor Networks," The International Journal of High Performance Computing Applications, Vol. 16, No. 3, August 2002.

[5] H. Chen and H. Q. Ye. Asymptotic optimality of balanced routing, 2010.[6] S. T. Maguluri, R. Srikant, and L. Ying. Stochastic models of load balancing and scheduling in cloud computing clusters. In *Proc. IEEE Infocom.*, pages 702–710, 2012.

[7] Raj Gaurav, "An Efficient Broker Cloud Management System" ACAI '11 Proceedings of the International Conference on Advances in Computing and Artificial Intelligence ACM New York, NY, USA ©2011 ISBN: 978-1-4503-0635-5

[8] Kulasekaran A. Sivakumar, Mahalingam Ramkumar: An Efficient Secure Route Discovery Protocol for DSR. GLOBECOM 2007: 458-463.

[9] P.Mell, T.Grance, The NIST definition of Cloud Computing , Version 15, National Institute of Standards and Technology, October 7, 2009..

[10] Zhao G, Liu, Tang Y.Sun,W.Zhang F, Ye X,Tang N (2009) Cloud Computing : A statistics aspect of users. In: First International Conference on Cloud Computing (CloudCom), Beijing, Springer,Berlin, Hiedelberg,pp 347 – 358.

[11]Cloud Security Alliance (2011) Security Guidance for critical areas of focus in cloud computing v3.0.

[12] Khalid .A (2010) Cloud Computing : Applying issues in Small Business In : International conference on signal acquisition and processing (ICSAP '10) pp 278 – 281.

[13]MatherT.Kumaraswamy,S.Latif (2009) Cloud Security and Privacy O'Reilly Media Inc,Sebastopol,CA.

[14] A. Manjeshwar and D. P. Agarwal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," In 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, April 2001.

[15] I. Foster, Y. Zhao, I. Raicu, and S. Lu. Cloud computing and grid computing 360-degree compared. In Grid Computing Environments Workshop, 2008. GCE'08, pages 1–10, 2008.

[16] David B. Johnson, David A. Maltz, and Yih-Chun Hu;`` The Dynamic Source Routing protocol for mobile ad hoc networks (DSR). Internet draft (work in progress)", IETF, April 2003.

[17] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, "Survivable WDM mesh networks," IEEE/OSA J. Lightwave Technol., vol. 21, no. 4, pp. 870–883, Apr. 2003.

[18] "Spectral grids for WDM applications: DWDM frequency grids," ITUT, Tech. Rep. G.694.1, June 2002.

[19] A. Nag, M. Tornatore, and B. Mukherjee, Fellow, Ieee, "Optical network design with mixed line rates and multiple modulation formats," IEEE/OSA J. Lightwave Technol., vol. 28, no. 4, pp. 466–475, Feb.2010. .