IOSR Journal of Engineering (IOSRJEN)
ISSN: 2250-3021 Volume 2, Issue 7(July 2012), PP 65-71
www.iosrjen.org

Efficient load Balancing in Cloud Computing using Fuzzy Logic

Srinivas Sethi[1], Anupama Sahu[2] Suvendu Kumar Jena[3]
[1, 2, 3] *Department of CSEA, IGIT Sarang, Odisha, India*

ABSTRACT : *Cloud computing is an expanding area in research and industry today, which involves virtualization, distributed computing, internet, software and web services. A cloud consists of several elements such as clients, data centers and distributed servers, internet and it includes fault tolerance, high availability, effectiveness, scalability, flexibility, reduced overhead for users, reduced cost of ownership, on demand services and etc. The services of cloud computing are becoming ubiquitous, and serve as the primary source of computing power for different applications like enterprises and personal computing applications. In this paper we introduced the novel load balancing algorithm using fuzzy logic in cloud computing, in which load balancing is a core and challenging issue in Cloud Computing. The processor speed and assigned load of Virtual Machine (VM) are used to balance the load in cloud computing through fuzzy logic.*
Keywords - *Cloud Computing; Load Balancing; Processing Time; Response Time;*

I. INTRODUCTION

Cloud computing [1][2] is an emerging commercial infrastructure and internet based cost-efficient computing paradigm in which information can be accessed from a web browser by customers for their requirement. It is a modality of computing characterized by on-demand availability of resources in a dynamic and scalable fashion, where the resource is used to represent infrastructure, platforms, software, services, or storage. It promises to eliminate the need for maintaining expensive computing facilities by companies and institutes alike. Cloud computing in a generally term is defined for anything that involves delivering hosted services over the Internet. It is based on the concept of shared computational, storage, network and application resources provided by a third party.

There are different kinds of clouds can be developed like private, public, and hybrid clouds [2]. In the private cloud the infrastructure is operated solely for an organization and managed by the organization or a third party and it may be exists on-premise or off-premise. In other words a public cloud sells services to anyone on the Internet and its infrastructures are made available to the public and are owned by an organization selling cloud services. A private cloud is a proprietary network or a data center that supplies hosted services to a limited number of people or group of people. When a service provider uses public cloud resources to create their private cloud, the result is called a virtual private cloud. The cloud infrastructures are shared by several organizations and supports a specific community that has shared concerns and it may be managed by the organizations or a third party and may be existed on-premise or off-premise. A hybrid cloud is a cloud computing environment which is combining the benefit of different types of cloud. An organization provides and manages some resources in-house and has others provided externally in hybrid cloud. Ideally, the hybrid cloud approach allows a business to take advantage of the scalability, efficiency, flexibility and cost-effectiveness.

The cloud provider is responsible to make the needed resources available on-demand to the cloud users and to manage its resources in an efficient way, so that the user needs can be meet [2]. The cloud computing is a model for enabling convenient, flexible, reliable on-demand network access to a shared pool of configurable computing resources, that can be rapidly provisioned and released with minimal management effort, which promotes availability and is composed of important characteristics like on-demand self-service, broad network access, resource pooling, rapid elasticity, measured services and efficiency and scalability; multi layered service abstraction like software as a service (SaaS), Platform as a service (PaaS) and infrastructure as a service (IaaS); and deployment models like private cloud, public cloud and hybrid cloud [2]-[4].

In the SaaS cloud model, the capability provided to the consumer is to use the provider's applications running on a cloud infrastructure and the applications are accessible from various client devices through a web browser. The consumer does not manage the underlying cloud infrastructure including network, servers, storage, operating systems and application capabilities, with the possible exception of limited user-specific application configuration settings. The vendor supplies the hardware infrastructure, the software product and interacts with the user through a front-end portal for the cloud users. The services can be anything from Web-based email to inventory control and database processing. The end-user is free to use the service from anywhere as the service provider hosts both the application and the data.

PaaS in the cloud computing environment is defined as a set of software and product development tools hosted on the provider's infrastructure. Developers create applications on the provider's platform over the

Internet. The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or consumer-acquired applications created using programming languages and tools supported by the provider. It has control over the deployed applications and possibly application hosting environment configurations. But the consumer does not manage the underlying cloud infrastructure. PaaS providers may use APIs, website portals or gateway software installed on the customer's computer.

IaaS in cloud computing aim at the hardware level and make the provisioning of customer resources such as servers, storage, connections and related tools. Building an IT infrastructure to provide IT services can be quite complex and companies are able to scale up or down their resources depending on the current workloads within minutes. The Data Center entity manages a number of host entities and the hosts are assigned to one or more virtual machine (VMs) based on a VM allocation policy that should be defined by the Cloud service provider. In IaaS cloud architecture, it is generally not known how the virtual machine will be used.

The capability provided to the consumer is to provision processing, server, storage, networks, and etc. It has control over operating systems, storage, deployed applications. The consumer is able to deploy and run arbitrary software, which can include operating systems and applications, but they does not manage the underlying cloud infrastructure. In the enterprise, cloud computing allows a company to pay for only as much capacity as is needed.

The rest of the paper is started with related work in section II, followed by load balancing in section III. Proposed work discussed in section IV and results are observed in section V. Finally conclusion is discussed in section VI.

II. RELATED WORKS

The authors in the paper [4] explores the concept of cloud computing, its advantages and disadvantages and describes several existing cloud computing platforms and discuss the results of quantitative experiments carried out using PlanetLab, a cloud computing platform as well. A two-phase scheduling algorithm under a three-level cloud computing network is a scheduling algorithm combines OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling algorithms that can utilize more better executing efficiency and maintain the load balancing of system[6].

In [3] the authors propose to find the best EFFICIENT cloud resource by Co-operative Power aware Scheduled Load Balancing solution to the Cloud load balancing problem. The authors designed the algorithm using the inherent efficiency of the centralized approach, energy efficient and the fault-tolerant nature of the distributed environment like Cloud

PALB [7], maintains the state of all compute nodes, and based on utilization percentages, decides the number of compute nodes that should be operating. It presents a load balancing approach to IaaS cloud architectures based on power as per local cloud computing architectures and it provides adequate availability to compute node resources while decreasing the overall power consumed by the local cloud.

Companies and institutions emphasize on cloud computing service and its application for the storage and analysis of very-large images, which has been implemented using multiple distributed and collaborative agents. A distinctive goal of this work is that data operations are adapted for working in a distributed mode by using different sub-images that can be stored and processed separately by different agents in the system, facilitating processing very-large images in a parallel manner [8].

In clouds, load balancing, as a method, is applied across different data centers to ensure the network availability by minimizing use of computer hardware, software failures and mitigating recourse limitations, where the availability of cloud systems is one of the main concerns of cloud computing [9]. In [10] the authors analyze the performance of cloud computing services for scientific computing workloads and quantify the presence in real scientific computing workloads of Many-Task Computing (MTC) users that, of users who employ loosely coupled applications comprises many tasks to achieve their scientific goals. They also perform an empirical evaluation of the performance of four commercial cloud computing services.

III. LOAD BALANCING

Load balancing is a process of reassigning the total load to the individual nodes of the collective system to the facilitate networks and resources to improve the response time of the job with maximum throughput in the system [13]. The important things in said load balancing are estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones to consider while developing such algorithm [14] . To improve the performance substantially, backup plan in case the system fails even partially, maintain the system stability,

accommodate future modification in the system are main goal of load balancing. Two main categories of load balancing algorithms are available; they are named as static and dynamic.

Static load balancing algorithms is an algorithm, which divide the traffic equivalently as round robin between available servers in which, the traffic on the servers will be disdained easily and consequently it will make the situation more imperfectly. It needs prior knowledge of the System. The weighted round robin another version of the round robin load is also defined to improve the critical challenges associated with round robin and each server has been assigned a weight. As per the highest weight they received more connections [15].

In dynamic load balancing the previous state or behavior of the system does not consider and it depends on the present state or behavior of the system. Decisions on load balancing are based on current state of the system with no prior knowledge. So it is better than static approach load balancing approach. Dynamic load balancing algorithms designated suitable weights on servers and by searching in whole network a lightest server preferred to balance the traffic. However, selecting an appropriate server needed real time communication with the networks, which leads extra traffic added on system [15].

Load balancing in the cloud computing based on standard load balancing but differs from classical thinking on load-balancing architecture and implementation by using commodity servers to perform the load balancing, which provides for new opportunities and economies of scale as well as presenting its own unique set of challenges. The load balancers served to promote availability of cloud resources and to promote performance [16]. In complex and large systems, there is a need for load balancing to simplify it in a cloud computing enviornment.

IV. PROPOSED WORK

A. A Fuzzy-based load balancing

In this paper we designed a new load balancing algorithm based on round robin in Virtual Machine (VM) environment of cloud computing in order to achieve better response time and processing time. The load balancing algorithm is done before it reaches the processing servers the job is scheduled based on various parameters like processor speed and assigned load of Virtual Machine (VM) and etc. It maintains the information in each VM and numbers of request currently allocated to VM of the system. It identify the least loaded machine, when a request come to allocate and it identified the first one if there are more than one least loaded machine. We tried to implement the new load balancing technique based on Fuzzy logic. Where the fuzzy logic is natural like language through which one can formulate their problem.

The advantages of fuzzy logic are easy to understand, flexible, tolerant of imprecise data and can model nonlinear functions of arbitrary complexity, and is used to approximate functions and can be used to model any continuous function or system [17][18]. Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic and the mapping provides a basis from which decisions can be made, or patterns recognized.

In our proposed work, the fuzzifier performs the fuzzification process that converts two types of input data like processor speed and assigned load of Virtual Machine (VM) and one output like balanced load which are needed in the inference system. In our proposed work, we considered the processor speed and load in virtual machine as two input parameters to make the better value to balance the load in cloud using fuzzy logic. These parameters are taking as inputs to the fuzzifier, which are used to measure the balanced load as the output.

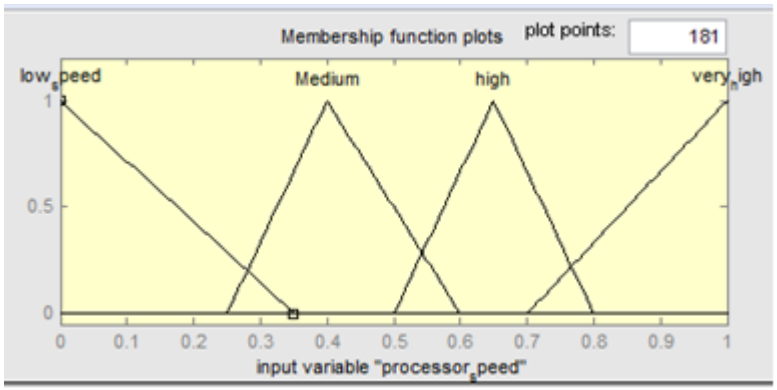

Fig-1: Membership input function of Processor Speed

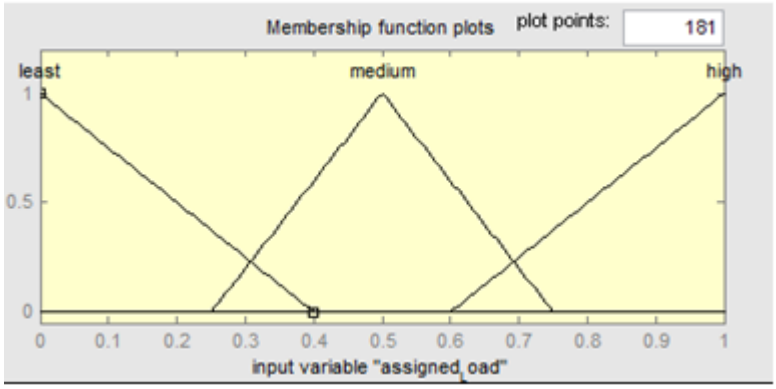

Fig.-2: Membership input function of Assigned Load

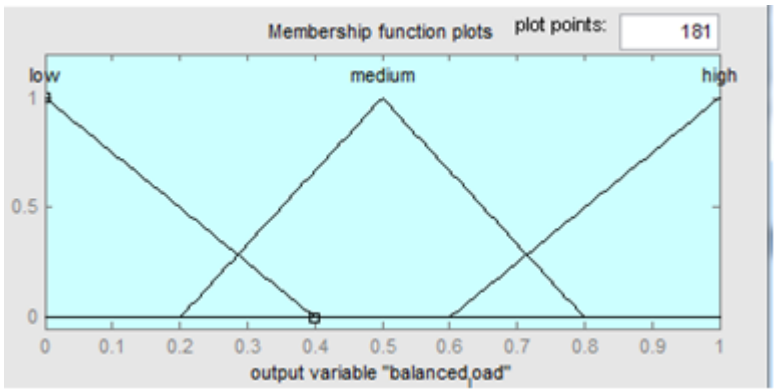

Fig.- 3: Membership output function of Balanced Load

1. If (processor_speed is low_speed) and (assigned_Load is least) then (balanced_load is medium) (1)
2. If (processor_speed is low_speed) and (assigned_Load is medium) then (balanced_load is low) (1)
3. If (processor_speed is low_speed) and (assigned_Load is high) then (balanced_load is low) (1)
4. If (processor_speed is Medium) and (assigned_Load is least) then (balanced_load is high) (1)
5. If (processor_speed is Medium) and (assigned_Load is medium) then (balanced_load is medium) (1)
6. If (processor_speed is Medium) and (assigned_Load is high) then (balanced_load is low) (1)
7. If (processor_speed is high) and (assigned_Load is least) then (balanced_load is high) (1)
8. If (processor_speed is high) and (assigned_Load is medium) then (balanced_load is medium) (1)
9. If (processor_speed is high) and (assigned_Load is high) then (balanced_load is medium) (1)
10. If (processor_speed is very_high) and (assigned_Load is least) then (balanced_load is high) (1)
11. If (processor_speed is very_high) and (assigned_Load is medium) then (balanced_load is high) (1)
12. If (processor_speed is very_high) and (assigned_Load is high) then (balanced_load is medium) (1)

Fig.-4: Use base rules

We used Fuzzy Inference System (FIS) to simulate human decision making based on the fuzzy control rules and the related input linguistic parameters. The low-high inference method is used to associate the outputs of the inferential rules [19][20]. Using the rule-based structure of fuzzy logic, a series of IF-THEN rules are defined for the output response given the input conditions, in which the rule is composed of a set of linguistic control rules and the accompanying control goals in the system. In our proposed work, there are 12 possible logical-product output response conclusions as shown in Fig.-4.

The Defuzzification is the process of conversion of fuzzy output set into a single number and the method used for the defuzzification is smallest of minimum (SOM). The aggregate of a fuzzy set includes a range of output values and be defuzzified in order to resolve a single output value from the fuzzy set. Defuzzifier adopts the aggregated linguistic values from the inferred fuzzy control action and generates a non-fuzzy control output, which represents the balanced load adapted to load conditions. The defuzzification method is employed to compute the membership function for the aggregated output [19][20].

The algorithm-1 is described to maintain the load in VM of cloud computing as follows.

Algorithm-1

```
Begin
   Connect_to_resource()
   L1
   If (resource found)
      Begin
         Calculate connection_string()
         Select fuzzy_connection()
         Return resource to requester
      End
   Else
      Begin
         If (Anymore resource available)
            Choose_next_resource()
            Go to L1
         Else
            Exit
      End
End
```

The proposed algorithm starts with request a connection to resource. It tests for availability of resource. It Calculate the connection strength if the resource found. Then select the connection, which is used to access the resource as per processor speed and load in virtual machine using fuzzy logic.

B. Experiment setup and parameter

Cloudsim[2] is a cloud simulator toolkit that enables modeling and simulation of Cloud computing systems and application provisioning environments. It supports both system and behavior modeling of Cloud system components such as data centers, virtual machines (VMs) and resource provisioning policies. Using the parameters mentioned in the table-1 of virtual machine and data centre, we have examined on cloudsim-3.0 a cloud computing simulator and evaluate the performance of proposed work using Average Response time, Data Center processing time.

Table-1

Parameter		Values
Virtual Machine	Image Size	10000
	Memory	1Gb
	Bandwidth	1000
Data Centre	Architecture	X86
	OS	Linux
	VMM	Xen
	Number of Machines	25
	Memory per Machine	2Gb
	Storage per Machine	100Gb
	Bandwidth per Machine	10000
	Number of processors per Machine	5
	Processor Speed	100MIPS
	VM Policy	Time Shared

V. RESULTS

In the result phase the main focus is to show the result, as the proposed load balancer Fuzzy based Round Robin (FRR) performs well, when comparing to the Round Robin Load Balancer by considering all instruction length per request. We have simulated the result by exploiting 25 machines, 5 numbers of processors per machine, and hundreds of jobs with the parameters mentioned as table-1.

We present some experimental results of the performance increase in the implemented service by minimize the data centre processing time and overall response time. Benefits of using fuzzy logic on round robin policy of load balancing are shown. Fig.-5 shows the data centre processing times are minimized with respect to all instruction length per request for Fuzzy based Round Robin (FRR) load balancer as compared to conventional Round Robin (RR) load balancer. We observed the efficiency of proposed load balancer FRR in terms of overall response time from Fig-6. It decreases the overall response time in all respect of data centre processing times as compared to RR. From these figure we observed that the FRR is better than RR, which is our objective.

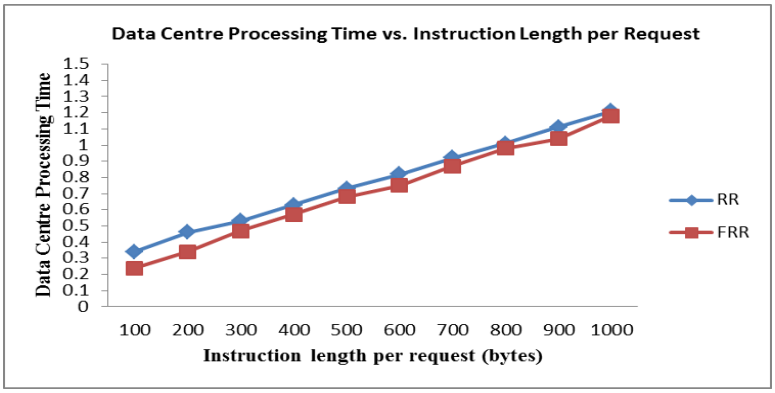

Fig-5: Data Centre Processing Time vs. Instruction Length per Request

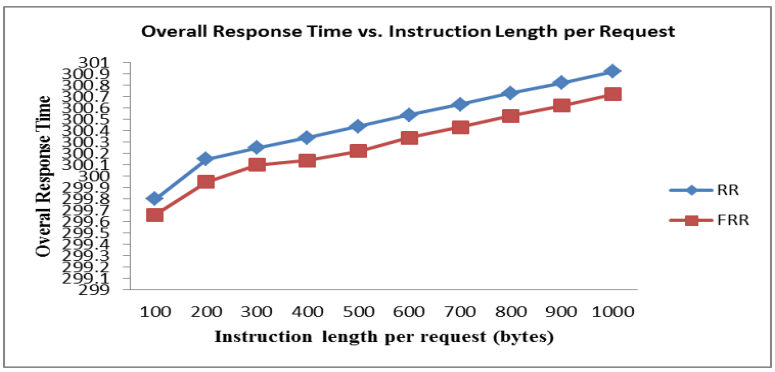

Fig6: Overall Response Time vs. Instruction Length per Request

VI. CONCLUSION

With the development of cloud computing as a paradigm in which scientific computing can done exclusively on resources leased only when needed as per users requirement from big data centers are faced with a new platform option. So, the initial target of clouds needs the scientific load balancing technique in the cloud computing environment. In this paper we proposed the new efficient load balancing technique using fuzzy logic based on Round Robin (RR) load balancing technique to obtain measurable improvements in resource utilization and availability of cloud-computing environment. The network structure or topology also required to take into consideration, when creating the logical rules for the load balancer. Two parameters named as the processor speed and assigned load of Virtual Machine (VM) of the system are jointly used to evaluate the balanced load on data centers of cloud computing environment through fuzzy logic. The results obtained with performance evaluation can balance the load with decreases the processing time as well as improvement of overall response time, which are leads to maximum use of resources. So, the obtained result shows the proposed Load Balancing algorithms (FRR) perform better than Round Robin (RR) Load balancer and it can be more appropriate in real life application efficient and effectively.

REFERENCES

[1] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, and Ivona Brandic, Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, Future Generation Computer Systems, Volume 25, Number 6, Pages: 599-616, ISSN: 0167-739X, Elsevier Science, Amsterdam, The Netherlands, June 2009.

[2] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, César A. F. De Rose, and Rajkumar Buyya, "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms", Software: Practice and Experience, Volume 41, Issue 1, pages 23–50, January 2011.

[3] T V R Anandarajan, M A Bhagyabini, "Co-operative scheduled Energy aware load-balancing technique for an efficient computational cloud", IJCSI, volume 8, issue 2, March 2011.

[4] D A Mansai, P NGO, "Understanding Cloud Computing: Experimentation and Capacity Planning", Proc. Computer Measurement Group Conf, Dallas, TX, Dec. 7-11, 2009.

[5] Menno Dobber, Rob van der Mei, and Ger Koole "Dynamic Load Balancing and Job Replication in a Global-Scale Grid Environment: A Comparison" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, vol. 20, no.2, pp 207, Feb 2009.

[6] Kuo-Qin Yan ; Wen-Pin Liao ; Shun-Sheng Wang , "Towards a Load Balancing in a three-level cloud computing network ", 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), 2010, Vol-1, pp.-108-113, 2010.

[7] Jeffrey M. Galloway, Karl L. Smith, Susan S. Vrbsky, "Power Aware Load Balancing for Cloud Computing", Proceedings of the World Congress on Engineering and Computer Science 2011 Vol I WCECS 2011, October 19-21, 2011.

[8] Raul´ Alonso-Calvo, Jose Crespo, Miguel Garc´ıa-Remesal, Alberto Anguita and Victor Maojo, "On distributing load in cloud computing: A real application for very-large image datasets", International Conference on Computational Science, ICCS 2010, pp.-2669-2677, 2010.

[9] Zenon Chaczko Venkatesh Mahadevan, Shahrzad Aslanzadeh and Christopher Mcdermid, "Availability and Load Balancing in Cloud Computing", 2011 International Conference on Computer and Software Modeling IPCSIT vol.14, IACSIT Press, Singapore, 2011.

[10] Alexandru Iosup, Member, IEEE, Simon Ostermann,Nezih Yigitbasi, Member, IEEE, Radu Prodan, Member, IEEE, Thomas Fahringer, Member, IEEE, and Dick Epema, Member, IEEE, "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing", IEEE TPDS, MANY-TASK COMPUTING, NOVEMBER 2010.

[11] Martin Randles, David Lamb, A. Taleb-Bendiab,"AComparative Study into Distributed Load Balancing Algorithms for Cloud Computing", IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, 20-23, pp. 551-556, April 2010.

[12] Meenakshi Sharma, Pankaj Sharma, Sandeep Sharma, "Efficient Load Balancing Algorithm in VM Cloud Environment", IJCST Vol. 3, Issue 1, pp.- 439-441, Jan. - March 2012

[13] R. Shimonski. "Windows 2000 & Windows Server 2003 Clustering and Load Balancing", Emeryville. McGraw-Hill Professional Publishing, CA, USA (2003), p 2, 2003.

[14] Ali M. Alakeel, "A Guide to Dynamic Load Balancing in Distributed Computer Systems", IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.6, June 2010.

[15] R. X. T. and X. F. Z, "A Load Balancing Strategy Based on the Combination of Static and Dynamic, in Database Technology and Applications (DBTA), 2010 2nd International Workshop 2010, pp. 1-4, 2010.

[16] IETF, 2010, Extensible Messaging and Presence Protocol (XMPP): CoreRFC3920, http://datatracker.ietf.org/doc/rfc3920/, viewed on 25th of Nov 2010.

[17] Y. F. Wong and W. C. Wong. "A fuzzy-decision-based routing protocol for mobile ad hoc networks" 10th IEEE International Conference on Network, pp. 317–322, 2002.

[18] G. V. S. Raju, G. Hernandez, and Q. Zou, "Quality of service routing in ad hoc networks", IEEE Wireless Communications and Networking Conference, Vol. 1, pp. 263 –265,2000.

[19] W. Pedrycz and F. Gomide, "An introduction to fuzzy sets: analysis and design", complex adaptive systems. MIT Press, 1998.

[20] J. J. Buckley, E. Eslami, and E. Esfandiar." An introduction to fuzzy logic and fuzzy sets" advances in soft computing, Physica Verlag, 2002.