International Journal of Engineering Research and Development
e-ISSN: 2278-067X, **p**-ISSN: 2278-800X, www.ijerd.com
Volume 7, Issue 7 (June 2013), PP. 54-60

A Review on Data Security, Accountability & Load Balancing in Cloud Computing

Mahesh Pavaskar[1], Vinayak Kankate[2], Harshal Khandre[3], Dr. B. B. Meshram[4]
[1,2,3]M. Tech Student, Computer Engg. Department,VJTI, Mumbai University
[4]Head of Computer Department, VJTI, Mumbai University

Abstract:- Cloud Computing provides new vision to the world. The approach of cloud computing is totally different from traditional system. In traditional system for any service, we need purchase, install, maintain, update by own. But in cloud environment we just pay for that service according to our usage basis & here no need to worry about purchase, install, maintenance and update. It is beneficial with respect to cost, flexibility, scalability. We require only good bandwidth network for better performance. Cloud provides different kinds of services like Software as a service, Platform as a service, Infrastructure as a service, Storage as a service etc. But still cloud not fully mature. Cloud computing facing following problems like Data Security, Monitoring and Latency Issues. In this paper we discus basic concept of cloud and review of how Security, Accountability & Load Balancing is achieved using different approaches like Encryption, Trusted Third Party Auditor (TPA) and effective resource utilization to reduce delay for produce output.

Keywords:- Cloud Computing, Data security, Accountability, Load Balancing.

I. INTRODUCTION

A. What is Cloud Computing

Cloud computing is the use of computing resources (hardware and software) which are available in a remote location and accessible over a network (typically the Internet). Users are able to buy these computing resources (including storage and computing power) as a utility, on demand. The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing is inclusive of virtualization to provide optimized resources, on-demand utilization, flexibility and scalability.

B. Why Cloud Computing

Some reasons why cloud computing is significantly more efficient:

a) Cost: Here we need to pay on Usage basis i.e according to Hrs/week/month/year. Instead of purchasing server we are using costly services according to our needs. So it is Affordable to small client use such services (SaaS, PaaS, IaaS).

b) Maintenance: Software‘s installed on Server. Client need not worry about installation, maintenance, updation & also need not to purchase license for each m/c.

c) Platform Independent i.e. OS (Win/Linux/Mac) Only requires high speed internet to use cloud services.

d) Flexibilit**y:** Cloud installations use virtualization and other techniques to separate the software from the characteristics of physical servers (some call this “abstraction of physical from virtual layers”).

C. Cloud Computing Service models

The services provided by cloud computing can be categorized into three service models, Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). These three models often abbreviated as the SPI Service framework (i.e. SPI is short for Software, Platform and Infrastructure) are the basis of all services provided by cloud computing:

• **Software as a Service (SaaS):**

In this model software is provided by the vendor over the net as a one-to-many model (single instance, multi-tenant architecture) as a substitute of the one-to-one typical model. Instead of users buying the software and installing it on their systems, they rent the software using pay-per-use, or subscription fee.[14] Thus the user exchanges the capital expense acquiring software licenses for operation expenses renting software usage. Since the application is provided over the net, usually the package includes the usage of the software itself and the utilization of the hardware it runs on, in addition to some level of support. Additional benefits of this model is centralized updating, so users don’t need to worry about patching and versioning. Examples of SaaS would be Google Docs and Salesforce.com customer relationship Management CRM software.

• Platform as a Service (PaaS):

The sophistication needed to create software that can run in the cloud entails the providers to create a development environment or platform on which these applications can be executed. The second service provided in the cloud is the utilization of the development environment itself. Users can create custom applications that target a certain platform, with tools offered by the platform provider. They then can deploy and run these applications on this platform, with full control over the applications and their configuration. Such applications may also be acquired from third parties. When using this service, users don't need, or even have the ability to manage the underlying cloud infrastructure, including servers, storage mediums and network configuration. The benefits of such a service are large, since startup companies and small teams can start developing and deploying their own software without the need to acquire servers and teams to manage them. Examples of PaaS would be Google's Apps Engine and Microsoft's Azure Platform.

• Infrastructure as a Service (IaaS):

In the third service, the users are given access to elements of the computing infrastructure itself. Using internet technologies, users can utilize the processing power, storage mediums and necessary networking components provided by the vendor. Users then can run arbitrary software and operating systems that best meets their requirements, with full control and management. This is much like traditional hosting services except when done in the cloud it is possible to scale the service to conform to the changing requirements, and to offer the pay per use model. This model is very similar to utility computing where users pay for the consumption of disk space, processing power, or bandwidth they use.Examples would be Amazon.com, EC2 and S3.

D. Types of Cloud

The cloud is divided into following category:Public (External), Private (Internal), Hybrid, Community cloud.[12]

Public: Public cloud applications, storage, and other resources are made available to the general public by a service provider. These services are free or offered on a pay-per-use model.

Private: Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party and services hosted internally or externally.

Hybrid: Hybrid cloud is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models.

Community cloud: It shares infrastructure between several organizations from a specific community with common concerns (security, accountability etc.), whether managed internally or by a third-party and hosted internally or externally.

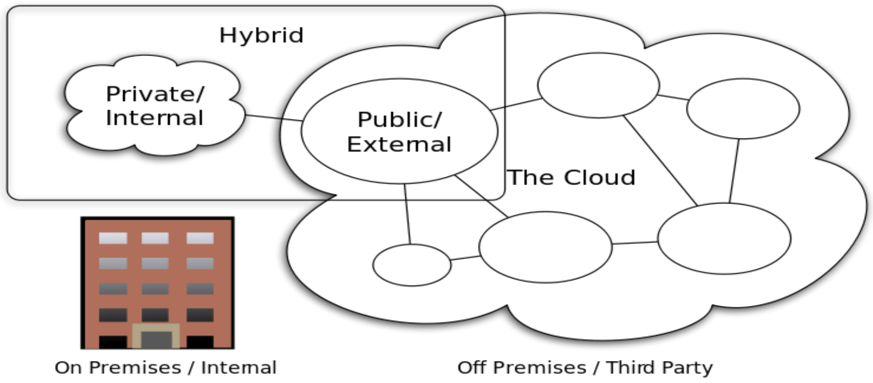


Figure 1: Cloud Computing Types

E. Cloud Overall Architecture

The cloud normally look like three tier architecture i.e SP, CP, CU.[12]

SP(Service Provider): The service provider provide services to the client. Services are provided according to client requirement. Example :- Google, Microsoft, Amazon etc

CP(Cloud Provider / Cloud Vendor): In the current market service provider and cloud provider are same. The function of both same. i.e provides services to the client.

CU(Cloud User): Cloud User may be small company or some organizations that needs some services from cloud provider.

II. LITERATURE SURVEY

[A]. Cloud Data Security

I. Information Security Policies

In Cloud computing technology there are a set of important policy issues, which include issues of privacy,security, anonymity, telecommunications capacity government surveillance, reliability, and liability, among others . But the most important between them is security and how cloud provider assures it. Well-known Gartner's seven security issues which cloud clients should advert as mentioned below [9]:

- Privileged user access: Sensitive data processed outside the enterprise brings with it an inherent level of risk because outsourced services bypass the "physical, logical and personnel controls" IT shops exert over in-house programs.
- Regulatory compliance: Customers are ultimately responsible for the security and integrity of their own data, even when it is held by a service provider [3]. Traditional service providers are subjected to external audits and security certifications.
- Data location: When clients use the cloud, they probably won't know exactly where their data are hosted. Distributed data storage is a usual manner of cloud providers that can cause lack of control and this is not good for customers who have their data in local machine before moving from local to cloud.
- Data segregation: Data in the cloud is typically in a shared environment alongside data from other customers. Encryption is effective but isn't a cure all. Encryption and decryption is a classic way to cover security issues but heretofore it couldn't ensure to provide perfect solution for it.
- Recovery: If a cloud provider broke or some problems cause failure in cloud sever what will happen to users' data? Can cloud provider restore data completely? Moreover clients prefer don't get permission to third-party companies to control their data. This issue can cause an impasse in security.
- Investigative support: Cloud services are especially difficult to investigate, because logging and data for multiple customers may be co-located and may also be spread across an ever-changing set of hosts and data centers.
- Long-term viability: Ideally, cloud computing provider will never go broke or get acquired by a larger company with maybe new policies. But clients must be sure their data will remain available even after such an event.

II. Network & Security Issues in Cloud Computing

There are different network issues occur in cloud computing some of which are discussed below:[12]

- Denial of Service
- Man in the Middle Attack
- Network Sniffing
- Port Scanning
- SQL Injection Attack
- Cross Site Scripting

Security issues of cloud computing are discussed below:[12]

- XML Signature Element Wrapping
- Browser Security
- Cloud Malware Injection Attack
- Flooding Attacks
- Data Protection
- Incomplete Data Deletion
- Locks in with vendor.

III. Solution to Cloud Security Problems

There are several traditional solutions to mitigate security problems that exist in the Internet environment, as a cloud infrastructure, but nature of cloud causes some security problem that they are especially exist in cloud environment [9][12]. In the other hand, there is also traditional countermeasure against popular Internet security problems that may be usable in cloud but some of them must be improved or changed to work effectively in it. A. Access Control Access control mechanisms are tools to ensure authorized user can access and to prevent unauthorized access to information systems. Therewith, formal procedures should be in place to control the allocation of access rights to information systems and services. Such mechanisms should cover all stages in the lifecycle of user access, from the initial registration of new users to the final de-registration of users who no longer require access to information systems and services. Special attention should be given, where appropriate, to the need to control the allocation of privileged access rights, which allow users to override system controls. The following are the six control statement should be considered ensuring proper access control management [9]:

1. Control access to information.
2. Manage user access rights.
3. Encourage good access practices.
4. Control access to network services.
5. Control access to operating systems.
6. Control access to applications and systems.

[B]. Cloud Data Accountability

The Cloud Information Accountability framework proposed in this work conducts automated logging and distributed auditing of relevant access performed by any entity, carried out at any point of time at any cloud service provider. It has two major components: logger and log harmonizer. The logger is strongly coupled with user's data (either single or multiple data items). Its main tasks include automatically logging access to data items that it contains, encrypting the log record using the public key of the content owner, and periodically sending them to the log harmonizer. It may also be configured to ensure that access and usage control policies associated with the data are honored. The log harmonizer is responsible for auditing. Being the trusted component, the log harmonizer generates the master key. It holds on to the decryption key for the IBE key pair, as it is responsible for decrypting the logs. Alternatively, the decryption can be carried out on the client end if the path between the log harmonizer and the client is not trusted. In this case, the harmonizer sends the key to the client in a secure key exchange. It supports two auditing strategies: push and pull. Under the push strategy, the log file is pushed back to the data owner periodically in an automated fashion. The pull mode is an on-demand approach, whereby the log file is obtained by the data owner as often as requested.

[C]. Load Balancing in Cloud

I. Load Balancing Parameter:

The performance of various load balancing algorithms is measured by the following parameters.

A. Overload Rejection: If Load Balancing is not possible additional overload rejection measures are needed. When the overload situation ends then first the overload rejection measures are stopped.After a short guard period Load Balancing is also closed down.

B. Fault Tolerant: This parameter gives that algorithm is able to tolerate tortuous faults or not. It enables an algorithm to continue operating properly in the event of some failure. If the performance of algorithm decreases, the decrease is proportional to the seriousness of the failure, even a small failure can cause total failure in load balancing.

C. Forecasting Accuracy: Forecasting is the degree of conformity of calculated results to its actual value that will be generated after execution. The static algorithms provide more accuracy than of dynamic algorithms as in former most assumptions are made during compile time and in later this is done during execution.

D. Stability: Stability can be characterized in terms of the delays in the transfer of information between processors and the gains in the load balancing algorithm by obtaining faster performance by a specified amount of time.

E. Centralized or Decentralized: Centralized schemes store globl information at a designated node. All sender or receiver nodes access the designated node to calculate the amount of load-transfers and also to check that tasks are to be sent to or received from. In a distributed load balancing, every node executes balancing separately. The idle nodes can obtain load during runtime from a shared global queue of processes.

F. Nature of Load Balancing Algorithms: Static load balancing assigns load to nodes probabilistically or deterministically without consideration of runtime events. It is generally impossible to make predictions of arrival times of loads and processing times required for future loads. On the other hand, in a dynamic load balancing the load distribution is made during run-time based on current processing rates and network condition. A DLB policy can use either local or global information.

G. Cooperative: This parameter gives that whether processors share information between them in making the process allocation decision other are not during execution. What this parameter defines is the extent of independence that each processor has in concluding that how should it can use its own resources. In the cooperative situation all processors have the accountability to carry out its own portion of the scheduling task, but all processors work together to achieve a goal of better efficiency. In the non-cooperative individual processors act as independent entities and arrive at decisions about the use of their resources without any effect of their decision on the rest of the system.

H. Process Migration: Process migration parameter provides when does a system decide to export a process? It decides whether to create it locally or create it on a remote processing element. The algorithm is capable to decide that it should make changes of load distribution during execution of process or not.

I. Resource Utilization: Resource utilization include automatic load balancing A distributed system may have unexpected number of processes that demand more processing power. If the algorithm is capable to utilize resources, they can be moved to under loaded processors more efficiently.

II. Types of Load Balancing Algorithm

Load balancing is a relatively new technique that facilitates networks and resources by providing a maximum throughput with minimum response time. Dividing the traffic between servers, data can be sent and received without major delay. Different kinds of algorithms are available that helps traffic loaded between available servers. A basic example of load balancing in our daily life can be related to websites. Without load balancing, users could experience delays, timeouts and possible long system responses. Load balancing solutions usually apply redundant servers which help a better distribution of the communication traffic so that the website availability is conclusively settled [14]. There are many different kinds of load balancing algorithms available, which can be categorized mainly into two groups. The following section will discuss these two main categories of load balancing algorithms.

a) Static Algorithms

In this method the performance [3] [14] of the processors is determined at the beginning of execution. Then depending upon their performance the work load is distributed in the start by the master processor. The slave processors calculate their allocated work and submit their result to the master. A task is always executed on the processor to which it is assigned that is static load balancing methods are non-preemptive. The goal of static load balancing method is to reduce the overall execution time of a concurrent program while minimizing the communication delays. A general disadvantage of all static schemes is that the final selection of a host for process allocation is made when the process is created and cannot be changed during process execution to make changes in the system load.

➢ Round Robin and Randomized Algorithms

In the round robin [5] processes are divided evenly between all processors. Each new process is assigned to new processor in round robin order. The process allocation order is maintained on each processor locally independent of allocations from remote processors. With equal workload round robin algorithm is expected to work well. Round Robin and Randomized schemes [6] work well with number of processes larger than number of processors. Advantage of Round Robin algorithm is that it does not require inter-process communication. Round Robin and Randomized algorithm both can attain the best performance among all load balancing algorithms for particular special purpose applications. In general Round Robin and Randomized are not expected to achieve good performance in general case.

➢ Central Manager Algorithm

In this algorithm [10], A central processor selects the host for new process. The minimally loaded processor depending on the overall load is selected when process is created. Load manager selects hosts for new processes so that the processor load confirms to same level as much as possible. From the on hand information on the system load state central load manager makes the load balancing judgment. This information is updated by remote processors, which send a message each time the load on them changes. This information can depend on waiting of parent's process of completion of its children's process, end of parallel execution The load manager makes load balancing decisions based on the system load information, allowing the best decision when of the process created. High degree of inter-process communication could make the bottleneck state. This algorithm is expected to perform better than the parallel applications, especially when dynamic activities are created by different hosts.

➢ Threshold Algorithm

According to this algorithm, the processes are assigned immediately upon creation to hosts. Hosts for new processes are selected locally without sending remote messages. Each processor keeps a private copy of the system's load. The load of a processor can characterize by one of the three levels: under loaded, medium and overloaded.

b) Dynamic Algorithms

It differs from static algorithms in that the work load is distributed among the processors at runtime. The master assigns new processes to the slaves based on the new information collected [2] [7]. Unlike static algorithms, dynamic algorithms allocate processes dynamically when one of the processors becomes under loaded. Instead, they are buffered in the queue on the main host and allocated dynamically upon requests from remote hosts.

- Central Queue Algorithm

Central Queue Algorithm [12] works on the principle of dynamic distribution. It stores new activities and unfulfilled requests as a cyclic FIFO queue on the main host. Each new activity arriving at the queue manager is inserted into the queue. Then, whenever a request for an activity is received by the queue manager, it removes the first activity from the queue and sends it to the requester. If there are no ready activities in the queue, the request is buffered, until a new activity is available. If a new activity arrives at the queue manager while there are unanswered requests in the queue, the first such request is removed from the queue and the new activity is assigned to it. When a processor load falls under the threshold, the local load manager sends a request for a new activity to the central load manager. The central load manager answers the request immediately if a ready activity is found in the process-request queue, or queues the request until a new activity arrives.

- Local Queue Algorithm

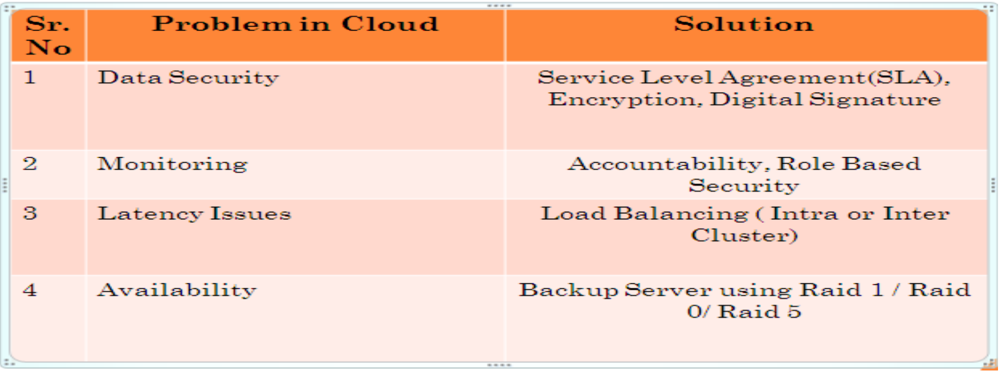
Main feature of this algorithm [12] is dynamic process World Academy of Science, Engineering and Technology 38 2008 270 migration support. The basic idea of the local queue algorithm is static allocation of all new processes with process migration initiated by a host when its load falls under threshold limit, is a user-defined parameter of the algorithm. The parameter defines the minimal number of ready processes the load manager attempts to provide on each processor. Initially, new processes created on the main host are allocated on all under loaded hosts. The number of parallel activities created by the first parallel construct on the main host is usually sufficient for allocation on all remote hosts. From then on, all the processes created on the main host and all other hosts are allocated locally. When the host gets under loaded, the local load manager attempts to get several processes from remote hosts. It randomly sends requests with the number of local ready processes to remote load managers. When a load manager receives such a request, it compares the local number of ready processes with the received number. If the former is greater than the latter, then some of the running processes are transferred to the requester and an affirmative confirmation with the number of processes transferred is returned.

c) Load Balancing in Cloud Computing

Cloud vendors are based on automatic load balancing services, which allowed entities to increase the number of CPUs or memories for their resources to scale with the increased demands [6]. This service is optional and depends on the entity's business needs. Therefore load balancers served two important needs, primarily to promote availability of cloud resources and secondarily to promote performance. According to the previous section Cloud computing will use the dynamic algorithm, which allows cloud entities to advertise their existence to presence servers and also provides a means of communication between interested parties. This solution has been implemented into the IETF’s RFC3920 - Extensible Messaging and Presence Protocol abbreviated as XMPP [7].

III. STATEMENT OF PROBLEM & POSSIBLE SOLUTIONS

Cloud computing has been envisioned as the next-generation architecture of IT enterprise. But current cloud computing facing following problems like Latency Issues, Availability, Data Security, Monitoring.In this project we can overcome these problem using Load balancing algorithm and using Third Party Auditor, key-pair exchange for trusted system



Sr. No	Problem in Cloud	Solution
1	Data Security	Service Level Agreement(SLA), Encryption, Digital Signature
2	Monitoring	Accountability, Role Based Security
3	Latency Issues	Load Balancing (Intra or Inter Cluster)
4	Availability	Backup Server using Raid 1 / Raid 0/ Raid 5

Table 1 : Issues in Cloud

IV. CONCLUSION

Cloud Computing is an applicable and interesting technology that introduce in the IT industry, it doesn't mean that all business IT needs to move to cloud. Moving towards cloud computing, we requires consider several parameters & most important of them is security. Using Encryption & Role based security we can provide security. Load balancing plays a vital role in case of clouds. Here we discussed some algorithms. The performance of algorithm depends on various parametes.

ACKNOWLEDGMENT

I render my sincere thanks to Dr. B. B. Meshram (Head Of Computer Department) VJTI, Mumbai for his guidance and encouragement.

REFERENCES

[1]. Smitha Sundareswaran, Anna C. Squicciarini, Member, IEEE, and Dan Lin, "Ensuring Distributed

[2]. Accountability for Data Sharing in the Cloud ", IEEE transactions on dependable and secure computing,

[3]. vol. 9, no. 4, july/august 2012.

[4]. Yan Zhu, Hongxin Hu, Gail-Joon Ahn, Senior Member, IEEE, Mengyang Yu, "Cooperative Provable

[5]. Data Possession for Integrity Verification in Multi-Cloud Storage ",IEEE transactions on parallel and

[6]. distributed systems

[7]. Cong Wang, Qian Wang, and Kui Ren & Wenjing Lou, "Ensuring Data Storage Security in Cloud Computing ",US National Science Foundation under grant CNS-0831963

[8]. Salvatore J. Stolfo, Malek Ben Salem, Angelos D. Keromytis, "Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud", Defense Advanced Research Projects Agency (DARPA)

[9]. Cong Wang, Kui Ren, and Jia Wang, "Secure and Practical Outsourcing of Linear Programming in Cloud Computing", IEEE TRANSACTIONS ON CLOUD COMPUTING April 10-15, 2011

[10]. Paul Marshall, Kate Keahey, Tim Freeman, "Improving Utilization of Infrastructure Clouds", IEEE/ACM Cloud Computing May 2011

[11]. M. Suresh Kumar and T. Purusothaman Computer Science & Engineering Department

[12]. Government College of Technology, Coimbator, India, "Penalty Based Heuristic Method using Continuous Double Auction in Computational Grid", European Journal of Scientific ResearchISSN 1450-216X Vol.70 No.4 (2012)

[13]. P. K. Suri and Manpreet Singh, Department of Computer Engineering M.M.Engineering College, M. M. University Mullana, Ambala, Haryana, India, "An Efficient Decentralized Load Balancing Algorithm For Grid", 978-1-4244-4791- IEEE 2010

[14]. Farzad Sabahi Faculty of Computer Engineering Azad University Iran, "Cloud Computing Security

[15]. Threats and Responses", 978-1-61284-486-IEEE 2011

[16]. Zhidong Shen and Qiang Tong School of Software, Northeastern University, "The Security of Cloud Computing System enabled by Trusted Computing Tech", 978-1-4244-6893- IEEE 2010

[17]. Tanveer Ahmed and Yogendra Singh , University School of Information Technology M.Tech CSE,GGSIPU, Dwarka New Delhi, India, "Analytic Study Of Load Balancing Techniques Using Tool Cloud Analyst", International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 2,Mar-Apr 2012

[18]. Sara Qaisar and Kausar Fiaz Khawaja, "Cloud Computing: Network/security threats and countermeasures", Institute of Interdisciplinary Business Research, Vol 3, no 9, Jan 2012

[19]. Zenon Chaczko, Venkatesh Mahadevan, Shahrzad Aslanzadeh and Christopher Mcdermid, University of Technology Sydney, Australia, "Availability and Load Balancing in Cloud Computing", International Con. on Computer and Software Modeling IPCSIT vol.14 (2011)

[20]. Pragati Priyadarshinee, Pragya Jain, "Load Balancing and Parallelism in Cloud Computing", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-5, June 2012

[21]. Qian Wang, Student Member, IEEE, Cong Wang, Student Member, IEEE, Kui Ren, Member, IEEE,Wenjing Lou, Senior Member, IEEE, and Jin Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", IEEE transactions on parallel and distributed systems, vol. 22, no. 5, may 2011

[22]. Sandeep Sharma, Sarabjit Singh, and Meenakshi Sharma, "Performance Analysis of Load Balancing Algorithms", World Academy of Science, Engineering and Technology 38 2008