Volume 7, Issue 5
October 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Challenges of Load Balancing Techniques in Grid Environment

[1,*]Richa Singh, [1]Arunendra Singh, [2]Pronaya Bhattacharya
[1]Pranveer Singh Institute of Technology, Kanpur,
Uttar Pradesh, INDIA
[2]Institute of Technology, Nirma University, Ahmedabad,
Gujarat, INDIA

E-mail: [1,*]rsrinet.876@gmail.com, [1]arun.sachan@gmail.com, [2]pronoya.bhattacharya@nirmauni.ac.in

ABSTRACT

Distributed system is the set of processing and computing elements (computer systems) connected over a communication medium. All computer systems have their own operating system, memory and processor. The jobs are executed concurrently on every computing systems in grid environment but some of the computing systems are heavily loaded and some are lightly loaded that cause a serious problem in terms of performance, communication delay and network latency. Lots of load balancing techniques in grid environment are suggested but all have some limitations. This paper discusses the basic concepts of load balancing in grid environment with their advantages and limitations along with future scope.

Keywords: *Distributed System, Load Balancing, Data Migration, Grid Computing System*

1. INTRODUCTION

Distributed computing system have a large number of nodes or a collection of various nodes which are connected within a communication network and utilize the workload effectively among all the nodes. Higher performance can be achieved with the help of effective utilization of the workload among the various nodes in the network.

All the nodes connected within a communication network works like a single node [1-3]. In the initial level every node have its own processing capabilities and speed and each node also initiates the load balancing on its own. To improve the load balancing and to make the task faster the workload should be evenly distributed to all nodes which are connected with each other in the grid computing architecture having their different processing capabilities. For this a uniform load balancing algorithm is required that can distribute workload evenly across the nodes. In grid computing architecture the load information coming from the different users which are connected in grid system are collected using Grid Load Management Tool(GLMT) which selects the resource according to the requested information[4-5]. GLMT is used to decide which resources are assigned to which task [6-7].

The load balancing system can be classified with the help of following points [8]:

1. Each node in grid system is monitored and managed.
2. In case of node failure share the information about resources and load so that the jobs are easily migrated from the other nodes.
3. Dynamically calculate the new load.
4. The current state of workload and other information are updated in real-time.

Grid computing[9] environment involves computing and processing resources and information system which are distributed everywhere and create a single point of contact. Grids are basically used for managing the workload among nodes to achieve better performance and higher utilization of resources. Managing the work load requires task scheduling and resource allocation. The main focus is to minimize the response time of a particular node and this will happen when the loads are evenly distributed on every node. There are various forms of load balancing algorithms which are discussed below:

Sharing the workload: In workload sharing there are two types of node one is busy and other is idle. The loads are shared with the idle node while the busy are already loaded.

Balancing the workload: In workload balancing the workloads are equally distributed among the nodes. There are various nodes in grid environment, some of them are lightly loaded and some of them are heavily loaded. In that case the workloads are transferred from heavily loaded to lightly loaded node for equal distribution of the workload [3] [7].

Levelling the workload: There should be no congestion on resources during the process of load distribution.

2. GRID COMPUTING ARCHITECTURE

In a grid computing architecture there are various groups consisting of several nodes. Each group have a lot of computing nodes connected via Local Area Network(LAN) connected to one switch. All these switches are connected to main switch via Wide Area Network(WAN). The groups are known as clusters as they form a mini cluster having some computing nodes. The basic concept of grid architecture is shown below which shows that how the nodes are connected with other nodes and switches.

Volume 7, Issue 5
October 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

In Figure 1 there are two parts in each group or cluster: the lower level and the upper level. The lower level are leafs or computing nodes which are connected to the switches. The upper level are the root node which are the switches in the group or cluster. They behave like a group manager and have responsibility to balance the load among the nodes.

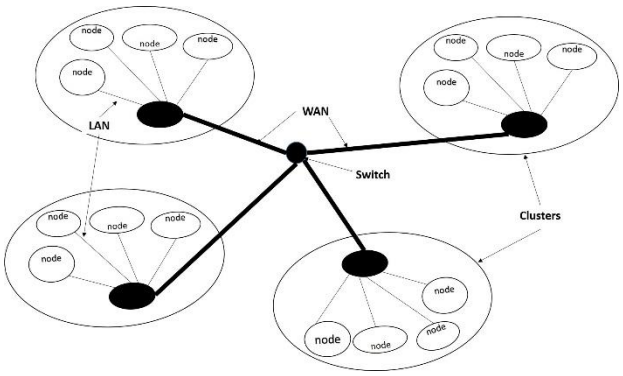

Fig 1: Architecture of Grid computing architecture

In this architecture there are two types of groups: one is inter group and other is intra group which are connected via a WAN. In intra group the group manager has the responsibility to balance the load among the nodes. The group manager calculates the workload of all the nodes and sends this information to other manager of the groups. If in case the group manager is not capable to balance the load among the nodes then at global level the load is balanced.

In inter group the grid finds out the working nodes and then start the load balancing. The group manager of inter group are responsible for load balancing among the nodes of different groups. Group manager send the information of workload to the respective group manager of intra group so that the distribution will be easy and also takes information of all the working nodes from their group manager.

3. CHARACTERISTICS OF GRID SYSTEM

There are various characteristics of a grid system discussed as follows:

1. **Large grid size**: A grid size may be quite large [10].
2. **Geographical distribution**: Resources may be located in different locations.
3. **Sharing of Resources**: In a grid environment the resources of other organization are used by the different users of different organizations. For cost reduction, to promote the efficiency non local resources can be used.
4. **Multiple administrations**: Each cluster have their own security and administrative polices. Resources can be used with the help of these policies. Different administration of heterogeneous clusters use the resources of other domain so in this case security is the major concern because security becomes more complex with different policies of different administrative domains.
5. **Resource coordination**: Resources are coordinated properly to aggregate computing capabilities [11].

4. CHALLENGES OF GRID SYSTEM

Various different hardware are used to create the grid and apart from this it is not possible for one person to manage these resources. To manage these resources effectively various different system administrators of different companies are used [12]. There are various challenges in grid computing some are listed below:

1. **Standards are not clear:** Various standards are used but same standards are not used by all the grid. All the operating systems use different standards [13-15]. In grid computing it is not possible to use different operating system in the same time at the same machine just because of different standards [16].
2. **Debate on what is grid and what is not:** The definition and the scope of grid computing is continuously changed so there is a lot of debate on grid computing [17-19].
3. **Difficulty in Grid Application Development:** The developments of grid applications are very difficult and expensive. The interfaces are designed according to the case with the help of computer scientists. Some scientists are not familiar with the parallel programming concept.
4. **Limited Area & Applications:** Grid computing has limited area and scope. For scientific applications more clusters can satisfy the application requirements. A specially designed super computer is required for special applications [20-21].
5. **Lack of grid enabled software:** Lots of effortis required to make a grid enabled software. Sometimes it is very difficult to reuse the software over grid environment.
6. **Sharing of Resources between Various Types of Services:** Different grid applications uses different grid platforms so it's very difficult to share resources among the various platforms.
7. **Management and Administration more challenging:** For any successful grid project management and administration also play an important role. Huge resources and services are distributed over the grid environment. It is a major challenge to manage and monitor all the resources and services distributed over the large geographical area [22].

5. NOTABLE TECHNIQUES FOR LOAD BALANCING IN GRID COMPUTING SYSTEM

Volume 7, Issue 5
October 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

A. Intra group vs. Inter group load balancing

Patni JC et al [23] suggested that Grid computing is the set of group's nodes from multiple administrative domains. In grid computing the two major key aspects are resource management and workload information management which provides a better service to grid environment users. There are various challenges in grid computing such as independency of resources, different type of load conditions, processing capabilities of the nodes overloading at computing nodes, resources are not properly utilized, large amount of computing nodes, nature of resources etc. To handle all these challenges in any kind of grid structure a distributed load balancing algorithm is proposed.

This proposed algorithm is further classified into two steps:

1. Execution time of the job is reduced.
2. Reduce the communication cost between two nodes when jobs are migrated in a grid architecture.

In this algorithm initially at the local level the loads are primarily balanced to reduce the communication cost.

The Proposed Load Balancing Grid Algorithm-

Input: A set of jobs $J=\{J_1,J_2,J_3,\ldots,J_n\}$ with their execution times in same ordering as $E=\{E_1,E_2,E_3,\ldots..E_n\}$. Also given is the set of processors $P=\{P_1,P_2,P_3,\ldots.P_n\}$ among a interprocessor network of nodes $N=\{N_1,N_2,\ldots..N_m\}$.

Output: An ordering schedule S such that the communication cost between two nodes is reduced when jobs are migrated in a grid architecture.

Assumptions: The algorithm proposes four functions-

1. Network_Toplogy(Nodes)-Initializes a communication topology among the nodes
2. Topological_Sort(*A*)- A is a list. A topological sort arranges the elements in a linear ordering or Direct Acyclic Graph based on an ordering relation.
3. add_node(*first_node,second_node,Network*)- places the two nodes in relative positions as defined by the network topology N.
4. BROADCAST(*W*)- broadcast a network instance to all nodes in the network.

GRID_ORDERING(*J,E,P,N*)
1. *for* all nodes *N*in the network
do initialize Network_Topology(*N*)
end*for*
2. *while**J* is not Empty
T←Toplogical_sort($J_1,J_2,J_3,\ldots,J_n$)
assign*T* in a priority queue *Q*
*while**Q* is not empty
W←dequeue(*T*)
X←dequeue(*T*)
for all processors*i*= 1 to *n*-1
do*in parallel*
C[i]←add_node(*W,X,N*)
for any two nodes N_k and N_l
L←compute_latency(N_k ,N_l)
computeL_{min}← min(*L*)
BROADCAST L_{min} for all *i*
end*for*
end *for*
S←BROADCAST($L_1,L_2,L_3\ldots L_{n-1}$)
end*while*
end*while*
3. The set S is the required ordering schedule.

Analysis: The cost of computing the load balancing in the distributed environment will depend on the sorting cost *S*of jobs and the propagationlatency cost *C* of the network divided among the processors operating in parallel and the cost of broadcasting information via messages in a distributed environment.

Advantages:

1. In grid computing environment the response time of the submitted jobs gets reduced.
2. Due to the job migration from one node to another the communication cost between two nodes are reduced.

Limitations:

1. **Time specific approach:** Each node sends workload information to other only in their own time limit.
2. **Locally Initiated Approach:** In this approach without informing the grid manager the nodes having higher load shares load to other nodes having lower load.

B. Develop integrated grid computing using power application- Karajan Workflow

Khannak R Al et al [24] proposed that a Distributed power system must ensure the power source is working in the desired allocated parameters. The parameters must work in the current power load and check must be enforced at regular time intervals. In non-real time system it gives some limitations such as failure of power system and increase and decrease of power supply. In this paper Grid computing is used for this type of distributed power system. The resources of grid computing is used in real time load balancing in distributed power system.

Advantages:

1. Heterogeneous method is used.
2. Integration of grid computing in power industries are technically and financially feasible.
3. Any changes depends on the employee's acceptance.

Limitations:

1. More processing time

C. Extenics based load balancing mechanism (ELBM)

Tao Der-Fu et al. suggested thatin a distributed system, load balancing [26] is the major issue just because of the system performance. To overcome this issue various load balancing algorithms are introduced. All the algorithms have some pros and cons. Four dynamic load balancing policies [27] are introduced named as ELBM (Extenics

based load balancing mechanism) for transfering the load from one node to another load instead of using fixed threshold value, create an adaptive threshold [28-29]value to determine whether the new job [30-31] is migrated from one node to another or not. The adaptive threshold value is calculated by relative function of average response time.

Apart from this policy, by-pass transfer method is used. In by-pass transfer method any job that needs to migrate can directly be transferred to the designated node instead of central node first. Due to this the communication [32] cost is reduced. To reduce the job arrival failure [33], a template queue is added at the mater node, to make the system more reliable. After simulation, the comparison of proposed mechanism with other algorithms such as RT and ALBCII algorithm gives a conclusion that the proposed mechanism evaluate better system performance as compared to others.

Advantages:

1. Reduced the communication cost
2. Make system more reliable.
3. Better performance achieved with this proposed mechanism.

Limitations:

1. It is a static method approach.

6. CONCLUSION

This paper discusses about the load balancing in grid computing environment. This paper begins with the grid computing architecture the basic concepts, characteristics and their challenges. The overview of all the load balancing techniques/ algorithms in grid environment are discussed. Various proposed algorithms and techniques for load balancing in grid environment are discussed with their advantages, disadvantages along with their future scope.

A proposed algorithm for dynamic balancing of load in grid environment is suggested. The overall conclusion of this paper is that real time load balancing in grid computing system is better as compared to static load balancing approach. Memory based load balancing in grid structure play an important role in future.

REFERENCES

[1] Den Burger, M, 2010,"Collective Receiver-Initiated Multicast for Grid Applications, *Parallel and Distributed Systems, IEEE Transactions* on (Volume: 22, Issue: 2), pp. 231-244.

[2] Kfatheen, S.Vaaheedha, MiM-MaM, 2015, "A new task scheduling algorithm for grid environment", *Computer Engineering and Applications* (ICACEA), Ghaziabad, India, pp. 695 – 699.

[3] FeiXu, iAware, 2013, "Making Live Migration of Virtual Machines Interference-Aware in the Cloud", *Computers, IEEE Transactions* on (Volume: 63, Issue: 12), pp. 3012-3025.

[4] Miguel L. Bote-Lorenzo, et al., 2004, "Grid Characteristics and Uses: a Grid Definition", Postproc. Of the *First European across Grids Conference (ACGf03), Springer- Verlag* LNCS 2970, Santiago de Compostela, Spain, pp. 291-298.

[5] ThiloKielmann, et al., 2006, "Scalability in Grid. PPT Core GRID", Bridging *Global Computing with Grid (BIGG)*, Sophia Antipolis, France.

[6] David De Roure, et al., 2005, "The Semantic Grid: Past, Present and Future", *Proceedings of the 2nd Annual European Semantic Web Conference (ESWC2005)*, Volume 93, Issue 3.

[7] Qin Zheng, 2013, "On the Design of Mutually Aware OptimalPricing and Load Balancing Strategiesfor Grid Computing Systems", Computers, *IEEE Transactions* on (Volume: 63, Issue: 7), pp. 1802-1811.

[8] Zhu Y., 2003, "A survey on grid scheduling systems", Technical report, Department of Computer Science, *Hong Kong University of science and Technology*.

[9] Grid Computing, URL-on https://en.wikipedia.org/wiki/Grid_computing(Accessed on November 2018).

[10] Scale Grid Computing Down to Size, URL-https://www.networkworld.com/article/2339444/software/scale-grid-computing-down-to-size.html, (Accessed on November 2018)

[11] Seyyed Mohsen Hashemi, 2Amid KhatibiBardsiri"Cloud Computing Vs. GridComputing" Dean of the Software Engineering and Artificial Intelligence Department , Science and Research Branch, Islamic Azad University, Tehran, IRAN.

[12] David Munoz Sanchez "Comparison between security solutions in *Cloud and Grid Computing*" Helsinki University of Technology.

[13] K. Czajkowski, D. F. Ferguson, I. Foster, J. Frey, S. Graham, I. Sedukhin, D. Snelling, S. Tuecke, and W. Vambenepe, The WS-Resource Framework, http://www.globus.org/wsrf/, (Accessed on November 2018).

[14] Global Grid Forum, URL-http://www.ggf.org,(Accessed on November 2018)

[15] Open Grid Services Architecture, URL-http://www.ggf.org/Public_Comment_Docs/Documents/draft-ggf-ogsa-specv1.pdf (Accessed on November 2018).

[16] S. Tuecke, Kzajkowski, I. Foster, J. Frey, S. Graham, C.Kesselman, D. Snelling, and P. Vanderbilt,February 17, 2003, "Open Grid Services Infrastructure".

[17] I. Foster, C. Kesselman, and S. Tuecke,2001, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", *International Journal of High Performance Computing Applications*, 15 (3), pp. 200-222.

[18] SETI@Home, URL-http://setiathome.ssl.berkeley.edu/, (Accessed on November 2018).

[19] Tony Blair Speech, Science matters, URL-http://www.number-10.gov.uk/output/Page1715.asp, (Accessed on November 2018).

[20] The Earth Simulator Center, URL-http://www.es.jamstec.go.jp/(Accessed on November 2018).

[21] The Blue Gene Project, URL-http://www.research.ibm.com/bluegene/(Accessed on November 2018).

[22] G. Yang, H. Jin, M. Li, N. Xiao, W. Li, Z. Wu, Y. Wu, and F. Tang,June 2004, "Grid Computing in China", *Journal of Grid Computing*, Vol.2, No.2, pp.193-206.

[23] Patni_JCand Singh MA, 4-5 September 2015 "Distributed load balancing model for grid computing environment" 2015 *1st International Conference on Next Generation Computing Technologies (NGCT-2015)*Dehradun, India, pp 123-126.

[24] R. Al-Khannak, B. Bitzer, 2007,"Load Balancing for Distributed and Integrated Power Systems using Grid Computing"pp 123-127, 1-4244-0632-3/07/$20.00 ©2007 IEEE.

[25] S. Sheng, K.K. Li, Z. Xiangjun, April 2004 ,"Grid Computing for Load Modeling," *IEEE International Conference on Electric Utility Deregulation, Restructuring, and Power Technologies.*

[26] Der-Fu Tao, and Liang-Teh Lee,February 2006, "An Extenics-based Load Balancing Mechanism in Distributed Computing Systems",*IJCSNS International Journal of Computer Science and Network Security*, Vol.6 No.2B, pp 70-76.

[27] Kyung-Soo Lim, Chong-Gun Kim, and DGun Kim, Aug, 1995, "Dynamic Load Balancing in Distributed Computer Systems with Star Topology," *Proceedings of IEEE 15th International Conference on Distributed Computing System*s, pp. 514-517.

[28] W. Cai,1983, "The extension set and incompatible problem," *Journal of Scientific Exploration*, vol. 1, pp. 81-93.

[29] Zeng Zeng and Veeravalli, B.,July 2004, "Rate-based and queue-based dynamic load balancing algorithms in distributed systems," *ICPADS*, pp. 349 –356.

[30] T. L. Casavant and J. G. Kuhl,Nov.1988 "Effective of Response and Stability on Scheduling in Distributed Computing Systems*," IEEE Transactions on Software Engineering*, Vol. 14, No. 11, pp. 1578-1588.

[31] Liudong Xing and Shrestha, A.,July 2005, "Distributed Computer Systems Reliability Considering Imperfect Coverage and Common-Cause Failures," *ICPADS*, pp. 453 - 457.

[32] H. Ashihara, N. Mizuguchi, and M. Maekawa,1991,"Reduction of Information Exchange for Adaptive Load Sharing in Distributed Systems," *ICPADS*, pp. 92-96.

[33] Singhal Mukesh et.al, TMH edition 2001 "Advanced Concepts in Operating System", ISBN-13:978-0-07-047268-6,pp (71-77).

AUTHOR PROFILES

Richa Singh is presently working as an Assistant Professor at Pranveer Singh Institute of Technology, Kanpur in Department of Information Technology. She received M.Tech. (CS) from Pranveer Singh Institute of Technology (PSIT), Kanpur (U.P.) in 2015. She received her B.Tech. (IT) from Prasad Institute of Technology (PIT), Jaunpur (U.P.) in 2011. Her area of interest is Distributed System, Software Engineering and Operating System.

Arunendra Singh is presently working as an Assistant Professor in Department of Information Technology at Pranveer Singh Institute of Technology, Kanpur. He is Pursuing his Ph.D. in Computer Science and Engineering from Dr. A. P. J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh. He received his M.Tech. from Motilal Nehru National Institute of Technology(MNNIT), Allahabad in 2011 and B.Tech. from Harcourt Butler Technological Institute(HBTI), Kanpur, Uttar Pradesh in 2005. He has published several research papers in highly reputed Journals and conferences like, Springer, IEEE, and ACM. His area of interest includes Network Security, Optical Communications, Computer Networks etc.

Pronaya Bhattacharya is presently working as an Assistant Professor in Department of Computer Science and Engineering at Institute of Technology, Nirma University, Ahmedabad, Gujarat. He is pursuing his PhD in Computer Science and Engineering from Dr. A. P. J Abdul Kalam Technical University, Lucknow, Uttar Pradesh. He has a teaching experience of more than 8 years as an Assistant Professor in various universities. He has authored papers in various reputed journals and conferences like IEEE and ACM. His areas of interest include Network Security, Machine Learning and Optical Communications.