Volume 3, Issue 12, December 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

# Providing Information Entree for Profound Dispersed Applications

**Prof. Ajanthaa Lakkshmanan, A. Abirami Shri, E. Aruna**

*School of Computing Science and Engineering*

*VIT University, Vellore, India.*

***Abstract— Vast distributed applications provides the flexible environment for the network used applications and the uncertain nature of the distributed systems for large applications makes the data access uncertain and unreliable .In this paper we propose a method that will make the system accessible to both the performance and also the availability .The vast number of the data used applications require the "accessibility" for the different job allocations .we present the accessibility-aware resource selection techniques that it is possible to select the nodes that have the efficient data access to the remote the data sources ,the local data observations that are collected from the nodes nearby nodes are sufficient to know the accessibility of that node. The distributed applications due to their large access of the resources the data sharing may be lead to the unexpected issues and also the network reliability also causes the several problems that leads to the nodes inaccessible to the required resources and in the scientific applications where the large set of nodes are being accessed by the vast applications and then he reliability of these system is very important and we provide in this system the accessibility aware resource selection technique that selects the nodes that have the node access and according to the application requirement.***

***Keywords— distributed applications, accessibility, nodes***

## I. INTRODUCTION

The vast distributed system provide the a platform for the network applications and this led to increase of the distributed applications and the peer to peer computing but the major challenge in the systems is the network uncertainness and the bandwidth availability . For example the project named "TONIC" that requires a bandwidth of the 367 Kbps and the "TONIC" hosts shows only the 146 Kbps and it will affect the entire application and the its performance will degrade

Recent Scientific applications are highly data intensive and also require the access to the large amount of the dispersed data and these applications that are data intensive and used in the different fields such as the weather prediction and the astronomy and the biomedical .The large number of scientists around the world access these applications to test and use them and in this situation the data is in Peta bytes and then the bandwidth connectivity is crucial and the efficiency depends on the data delivery on different location of the data and point of access and it needed not only consider computational capability but also the data accessibility of the computational nodes that require the data objects.

Data Availability is considered as the crucial part in these systems and the availability is the primarily uses at the Server side metric that will restrict the client side accessibility of the data .for example that a data part may be with 98.99 percent but if we access it in real it is not accessible due to the misconfiguration and the network related problems even this file is available to the two clients and one have the bad connection to file server and resulting in the more download time ,hence it is necessary to consider the metric of the "data accessibility ".The main problem is that the accessibility of the each and every client nodes in the heavy connected distributed systems and this is made more difficult by the dynamics of the WAN(wide area networks) and the cost of the information gathering is also increases .For example suppose a node with the weak access link will slow access at all the times and by knowing the access information of the neighbours, it can be useful for probing.

## II. THE ACCESSIBILITY USED FOR RESOURCE SELECTION

In this paper we present a system that shows the data accessibility by calculating for the neighbour nodes .The System consists of the network of a compute nodes that will provide the computational resources for different applications and data nodes which require for the computation .

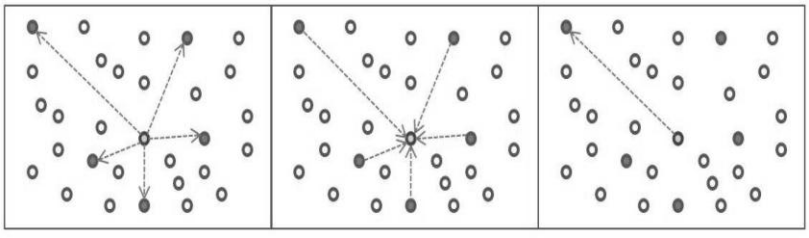



Fig. 1 Accessibility of the resource selection

The data objects we can assume them as the files or records and the data nodes and the compute are connected in the overlay structure. The over-lay structure that is constructed by the typical over-lay structure and the system. The Resource selection process is that it will access by the neighbours and initiators of the set of the candidates nodes which are chosen and they will give the candidates of first of the queries and it will give the all the information that needed for the access of the nodes and to do the job and it will also calculate the impact of the data access and the performance of the job which is executed

The Algorithm:

$$H_s : C \longrightarrow c_m \text{ such that}$$

$$accessibility_{c_m}(J_i) = \min_{n=1,..,|C|}(accessibility_{c_n}(J_i))$$

The Accessibility Resource selection "∞" which used for the candidates
The Down speed calculation:

$$\overline{DownSpeed_h} = \frac{1}{|\mathcal{H}_h|} \cdot \sum_{i=1}^{|\mathcal{H}_h|} \frac{\mathcal{H}_h^i.size}{\mathcal{H}_h^i.elapse}$$

The calculation of Self Estimation:

$$SelfEstim_h(o) = \delta \cdot \frac{size(o)}{\overline{DownSpeed_h}}$$

Where

$$\delta = \frac{distance_h(server(o))}{\overline{Distance_h}}$$

Here the "size (o)" means the size of the object "o"
And the "server (o)" means the server for the object o, and the "distance (b)" means the distance between the nodes a and b.

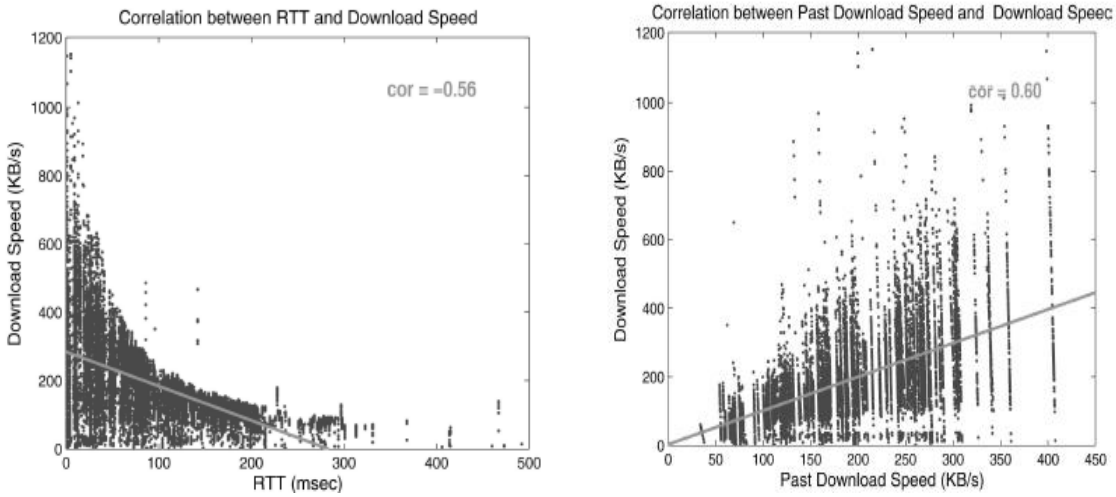



Fig. 2 The accessibility and the parameters

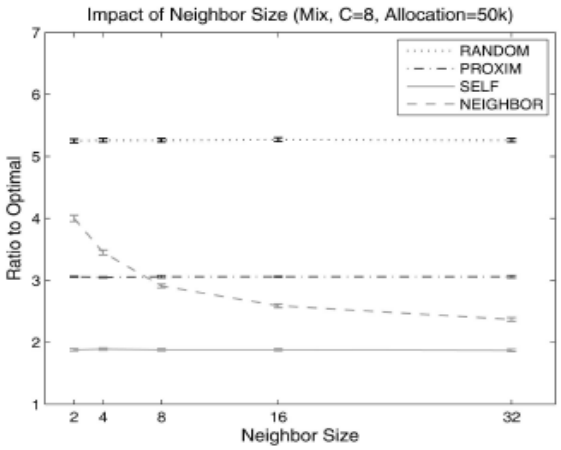
*A. The neighbour estimation*

The self –estimation which uses the nodes observations which is used to estimate the accessibility of the data object and have the accuracy of the estimation .The server location may become overhead and it also to latency for the resource of selection and to avoid the problems in this system we present the estimation as a new approach

*B. Impact of the neighbour size*

We have to consider the impact of the neighbour size by using "NEIGHBOUR" and the selection techniques are applied so that the neighbour size will improve the performance.

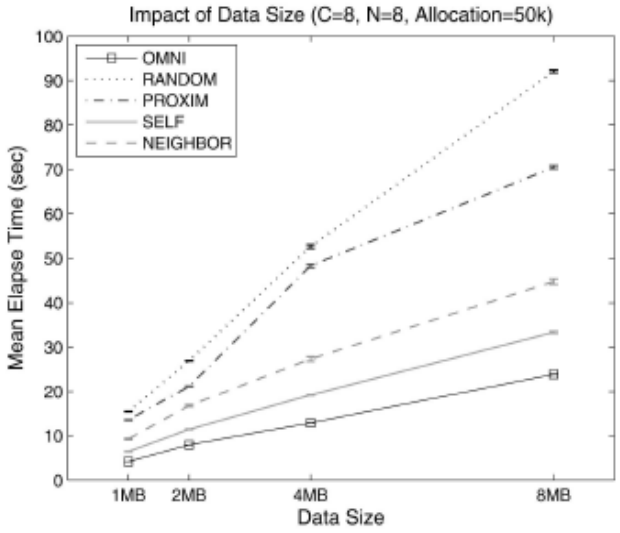



*C. Impact of the candidate size*

We have now consider the impact of the candidate size before the allocation of a job and its calculated with respect to the "candidate size", the "NEIGHBOUR" follows the "SELF" and this will also improve the performance.

*D. Impact of the data size*

There are many sizes of the data to be handled and it also depends on the application and also work for the selection techniques for different data sizes and we will do the individual traces than the mixed traces ,this ensures that the proposed system works well for the data intensive applications.




.

## III. SYSTEM ANALYSIS

*A. Scope of the project*

In scientific applications are data is intensive and it will require the some amount of data in this system we are accommodate with data intensive applications loosely coupled in distributed systems. It is very important to consider not only compute the capability we can also consider the data accessibility of the computational nodes, it will require for the objects of the data. Here we are considering the both performance and availability. In this application it is essential to not only consider the computation node power but also we consider the accessibility of job allocations. In this project, we are present that accessibility aware resource selection techniques, in which it is may be possible to choose the nodes that are having data access efficient for the remote data sources.

## IV. MODULE DESCRIPTION

*A. Designing the system model*

In our model contains the network to compute node ,this will provide resources for computation will execute the job applications and data nodes will store the objects of the data this will require for estimation our project. In our project data objects can be files or any other type of data representation but we are assuming that both the compute nodes and data nodes are connected like overlay structure. Here we are not assuming any particular type of system for this overlay. The system will be constructed like by using any of the typical network architecture like structured and unstructured or any other type of techniques but in this system we assumed that system will provides build in functions for storing the objects and retrieving the other objects and access to any node over the system. In this system each node may be data node or compute node or both.

*B. Checking accessibility*

In this system at first the initiator node will asks the accessibility estimation to all the candidate nodes. The candidate node will estimate accessibility to the initiator. The initiator node will select the compute node from the set of all candidates.

*C. Allocate job to the best node*

If once the initiator will select the node after that the job will be transferred to the one of the selected node. This node will be called as a worker. The worker is downloads the objects of the data require for the job from network after that implement the calculation whenever if the node execution is over, this worker will returned the whole result to the initiator.

*D. Existing system*

In existing system the ability to large scale systems are having the cycles for geographical distributed systems, this was growing in several projects like home projects.The main disadvantage of existing system is the network is unpredictable in nature and bandwidth also very limited for data, more cost and not scalable.

*E. Proposed system*

The system contains of networks for compute nodes that gives the resources for the computational nodes and this computational nodes executing the jobs for the applications and data nodes store the objects of the data ,this will require for computation. Actually, in our context files can be represented by the data objects and database records or any other type of representation of the data.

In this model we are assumed that compute nodes and data nodes both are connected like overlay structure .here we are not assume any particular organization type overlay .this system will be constructed by using any of the network architectures such as structured and unstructured or any other type of techniques.

However in this system will give build in functions for storing and retrieve of objects. This will be distributed and access from any of the other nodes over the system .in this network the nodes may be data nodes and compute node or both.

***Advantages:***

1. This proposed system is to give the good performance and high availability
2. In this system will provide the build in functions such for storing the objects and retrieving the objects this will gives distribute and access by any node over the system.
3. This will be cost effective and scalable in nature.

## V. SYSTEM DESIGN

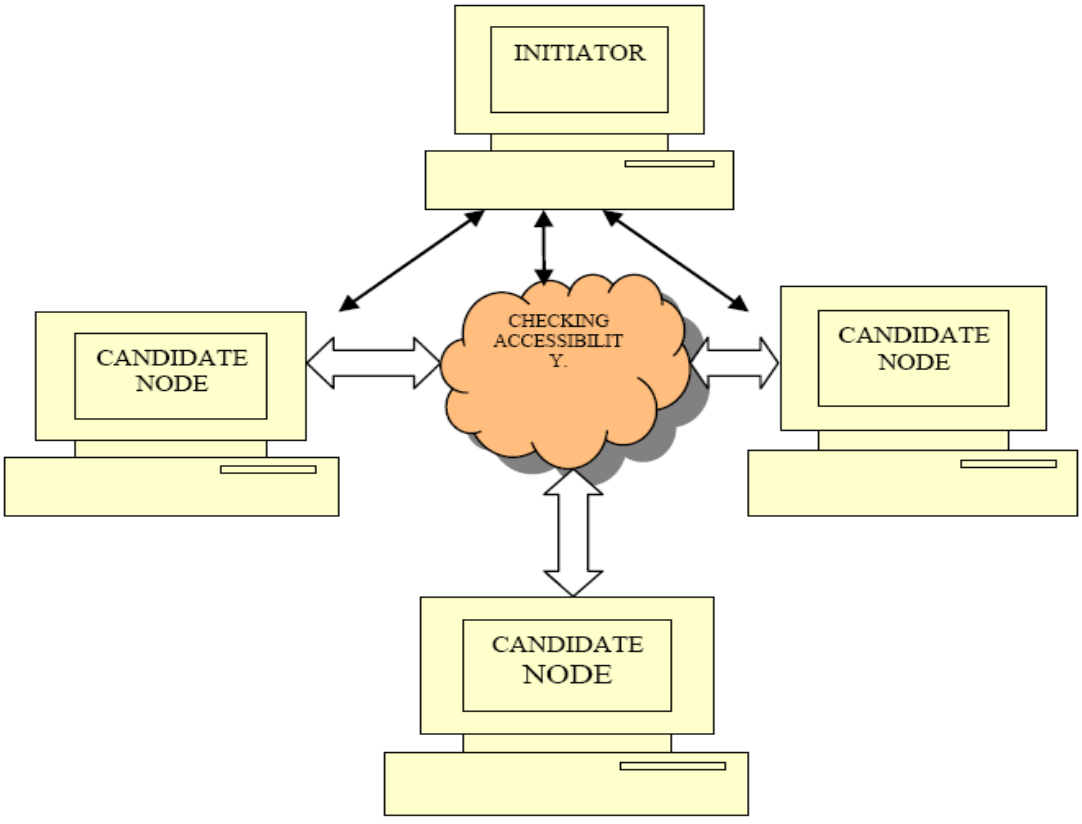



Fig. 3 First Level

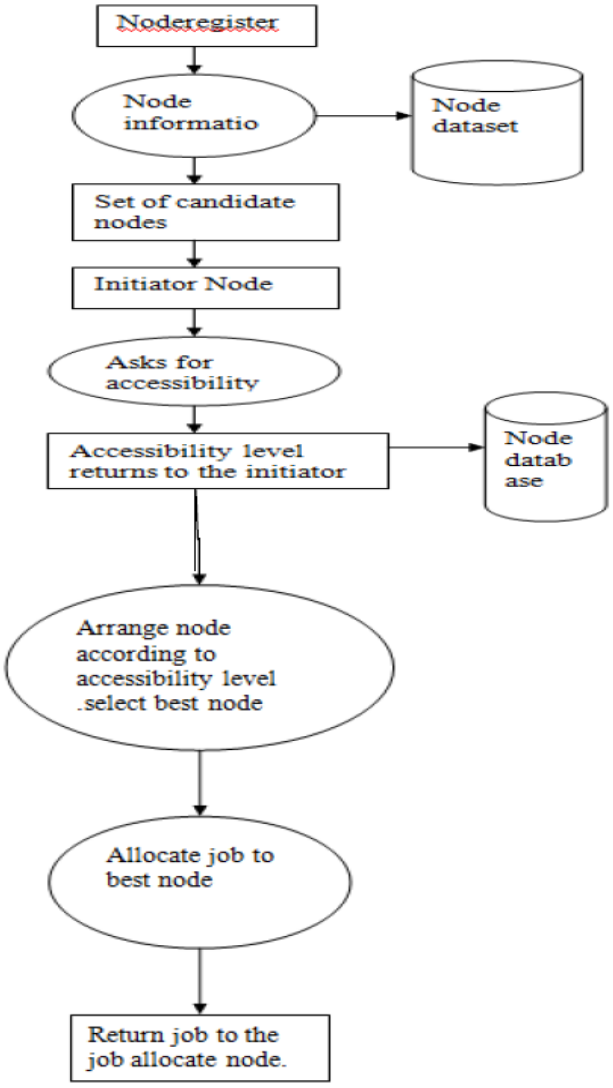
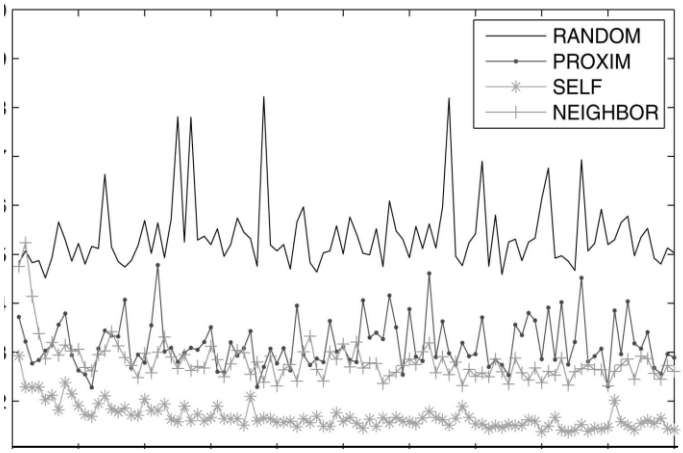



Fig.4 Second Level

## VI. EXPERIMENTAL ANALYSIS

In experimental analysis we deploy a network called pastry. We are distributing the objects of different sizes over the network .we are generating a set of Random queries that the nodes are to be performing the downloading objects. In this each object are having its own key. Here we are evaluating the selecting the resource , which we are giving the inputs to the ping –maps and the collective Downloading Traces and the results will be displayed according to the selecting the algorithms . At first the nodes are connected like randomly to one another with a fixed Neighbour –Size without considering the topology or locality. In the experimental analysis we compare the random selection and latency based selection and self based selection and neighbour selection.



## VII. CONCLUSION

Actually accessibility is very critical if the data intensive applications are increasing in loosely coupled distributed systems. This type of applications is requiring resource selection due to the unpredictable connection and bandwidth .In our methods we rely on historic and local observation, in this network overhead is tolerable. Compared to the conventional techniques we are nearer to the minimal selection. We also have seen that methods are uniquely out-performing conventional methods in the replicated environment. in this project we are focusing on the performance metric after that we are focusing on availability dynamically, there is no methods on the estimation of the neighbour on geographic or topological similarities are to be utilize that nodes are coming from the all the nodes. In above project, we explore scalable and decentralized and resource will be selected efficiently based on the accessibility. In this project our techniques are provide that accurate meaningful rank in order of based on the node accessibility.

**ACKNOWLEDGMENT**

The authors would like to thank the School of Computer Science and Engineering, VIT University, for giving them the opportunity to carry out this project and also for providing them with the requisite resources and infrastructure for carrying out the research.

**REFERENCES**

[1] D.P. Anderson and G. Fedak, "The Computational and Storage Potential of Volunteer Computing," Proc. IEEE Int'l Symp. Cluster Computing and the Grid (CCGRID '06), pp. 73-80, 2006.

[2] A. Haeberlen, A. Mislove, and P. Druschel, "Glacier: Highly Durable, Decentralized Storage Despite Massive Correlated Failures,"Proc. Symp. Networked Systems Design and Implementation (NSDI '05), May 2005.

[3] J. Kubiatowicz, D. Bindel, Y. Chen, P. Eaton, D. Geels, R. Gummadi, S. Rhea, H. Weatherspoon, W. Weimer, C. Wells, and B. Zhao, "Oceanstore: An Architecture for Global-Scale Persistent Storage," Proc. ACM Int'l Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS '07), Nov. 2000.

[4] R. Bhagwan, K. Tati, Y.-C. Cheng, S. Savage, and G.M. Voelker, "Total Recall: System Support for Automated Availability Management," Proc. Symp. Networked Systems Design and Implementation (NSDI '04), p. 25, 2004.

[5] A. Chien, B. Calder, S. Elbert, and K. Bhatia, "Entropia: Architecture and Performance of an Enterprise Desktop Grid System," J. Parallel and Distributed Computing, vol. 63, no. 5 pp. 597- 610, 2003.

[6] D. Kondo, A.A. Chien, and H. Casanova, "Resource Management for Rapid Application Turnaround on Enterprise Desktop Grids," Proc. ACM/IEEE Conf. Supercomputing (SC '04), p. 17, 2004.

[7] J.-S. Kim, B. Nam, P. Keleher, M. Marsh, B. Bhattacharjee, and A. Sussman, "Resource Discovery Techniques in Distributed Desktop Grid Environments," Proc. IEEE/ACM Int'l Conf. Grid Computing (GRID '06), Sept. 2006.

[8] D. Zhou and V. Lo, "Cluster Computing on the Fly: Resource Discovery in a Cycle Sharing Peer-to-Peer System," Proc. IEEE Int'lSymp. Cluster Computing and the Grid (CCGRID '04), pp. 66-73, 2004.

[9] D.P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D.Werthimer, "Seti@home: An Experiment in Public-Resource Computing," Comm. ACM, vol. 45, no. 11, pp. 56-61, 2002.

[10] "Search for Extraterrestrial Intelligence Project,"http://setiathome.berkeley.edu, 2009.

[11] "BOINC: Berkeley Open Infrastructure for Network Computing," http://boinc.berkeley.edu/, 2009.

[12] "PPDG: Particle Physics Data Grid," http://www.ppdg.net, 2009.

[13] N. Massey, T. Aina, M. Allen, C. Christensen, D. Frame, D.Goodman, J. Kettleborough, A. Martin, S. Pascoe, and D. Stainforth, "Data Access and Analysis with Distributed Federated Data Servers in climate prediction.net," Advances in Geosciences, vol.8, pp. 49-56, June 2006.

[14] G.B. Berriman, A.C. Laity, J.C. Good, J.C. Jacob, D.S. Katz, E. Deelman, G. Singh, M.-H. Su, and T.A. Prince,"Montage: The Architecture and Scientific Applications of a National Virtual Observatory Service for Computing Astronomical Image Mosaics," Proc. Earth Sciences Technology Conf., 2006