ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 8, October 2012

Perspective Study on Resource Information Aggregation in Computational Grid

R. Venkatesan, M. Manicka Raja

Abstract

Grid Computing is a mechanism of distributed computing to coordinate and share the networked resources, data domains, storage devices and processing power across geographically distributed locations. In the computational grid network, information services are basically providing information of infrastructure resources. The Information includes configuration of resources, policies, agreements are managed by both higher level and lower level schedulers. Resource information aggregation methods are used to reduce the quantity of information transferred between the networked resources. Through information aggregation, the characteristics of resources are summarized and then sent to the scheduling processes. This survey paper discusses about various resource information aggregation models of grid networks.

Key Terms – Grid Computing, Information aggregation, Resources, Schedulers.

I. Introduction

Grid is a distributed environment that makes it probable to share heterogeneous, loosely combined IT resources through organisations and geographical locations [1]. Grid network systems link computing resources together in an approach that contracts someone use one computer to access and force the collected power of all the computer machines in the environment [2]. Grid computing environments are the outcome of autonomic provisioning of a multitude of resources and capabilities, typically demonstrating increased computing resource utilization, access to specialized computer systems, cost sharing, and improved management capabilities. Grid Computing performs various applications in life sciences, financial analysis, research collaboration, engineering design, collaborative games, bio-informatics and etc [3].

The computational Grid [4] systems have higher computational capacity to perform complicated calculations. Due to the large amount of resources existing in computational grid networks, it is difficult to manage the resource information. Using information aggregation the characteristics of grid resources are grouped and then sent to the higher level scheduling mechanisms. The scheduler makes its choices depending on the information collected my resource monitoring systems [5]. The Grid model consists of two stages of schedulers. First the central scheduler receives the request from its higher level mechanism and decides which resource domain is suitable for executing the computation. In second level, the domain scheduler selects an appropriate resource site to solve the computational problem. With the use of information aggregation resource associated information are summarized and conveyed between these two level schedulers. This reduces the quantity of information exchanges among the grid machines [6].

This study paper has been organized as follows. Section II presents various resource information aggregation methods. Section III provides an analysis and parameter wise comparison among all the surveyed papers. Section IV concludes the survey and deals with future works.

II. Various resource information aggregation methods in grid computing

(a) Domain based resource information aggregation [7]:

Description:

Through this method the characteristics of resources are consistently signified and the amounts of information exchanged between the grid resources are reduced. The domain monitor collects the information from its resource sites and computes aggregation matrix. This matrix is sent to central monitor to make scheduling decisions.

(i) Algorithm 1 provides phases for the information collection and aggregation pattern.

(ii) Algorithm 2 presents the scheduling scheme that uses the aggregated information matrix by central scheduler. In this method some of aggregated parameters such as min, max, additive are used. Using Single point and Intra-Domain aggregation schemes information matrix are evaluated.

Merits:

(i) The job of the Central monitor is divided by Domain monitors. Therefore monitoring of resources can be easy.

(ii) Minimizing clumsy data transfer.

Demerits:

(i) Maintaining the Resource information matrix in Central monitor is a difficult one. Therefore efficient data structure is needed.

(b) Information aggregation through BestBrokerRank policy [8]:

Description:

In this model, all clusters or domains adopt a public data aggregation model that allows both the encapsulation and sharing of its resources and scheduling information. Every domain has a grid broker that acts as a firewall to that domain. Through best broker rank policy, the central scheduler selects an appropriate domain to execute the task. The two aggregation algorithm discussed here are,

(i) SIMPLE aggregation algorithm - The aggregation parameters used here are Processor type, OS type and File type. Some operators such as count, total are used to aggregate the parameters.

(ii) CATEGORIZED aggregation algorithm – It enables subcategories to increase the accuracy of the resource information and fix threshold values for aggregation.

Merits:

(i) High accuracy of information is achieved.

(ii) Latency time and communication bandwidth are saved.

(iii) Highly scalable to resource information size and processing time.

Demerits:

(i) Execution time is longer.

(ii) Resource information size is larger.

(c) Data Consolidation and Information Aggregation [9]:

Description:

In Grid networks, sometimes the scheduler needs more than one piece of data to schedule the resources. Here data consolidation method introduces to transfer group of data among the resource sites and data replicas. Four types of schemes are familiarized to summarize the resource related information.

(i) Random-Random scheme – It randomly choses DC sites and data replicas to schedule the tasks using Dijkstra algorithm.

(ii) Consolidation-Cost Scheme – It selects the data replicas and the DC which has minimal amount of DC time.

(iii) Execution-Cost Scheme – It selects the data replicas and the DC which has minimal amount of task execution time.

(iv) Total-Cost Scheme – It selects the replicas and the DC site which has minimal amount of total task delay. This delay includes the time needed for transferring the datasets to the DC site and the task's execution time using Dijkstra algorithm.

Merits:

(i) Grouping the domains can be easy.

(ii) Reduces network load.

Demerits:

(i) More no. of sites is used. Therefore implementation of routing algorithm can be a difficult one.

(ii) Every node in a network has more space queues for storing transmission data. Data maintenance is difficult.

(d) Aggregating Service with Failure Detector [10]:

Description:

This method introduces fault detection architecture called HYDRA to monitor the process failures in distributed systems. The aggregator component of HYDRA publicizes the characteristics of resources which were retrieved from a GIS (Grid Information System) to the grid schedulers. If a user enquiries about specific process, the aggregator sends its aggregated information. If there is no information about its status, the resource aggregator forwards the query to the GIS, which is related with the process.

Merits:

(i) Network failures can be detected easily.

(ii) High availability.

Demerits:

(i) Complicated architecture.

(ii) Low protocol security.

(e) Distributed Indexing based aggregation scheme [11]:

Description:

This method proposes a Multi-Attribute Addressable Network (MAAN) for resource indexing and a Distributed Aggregation Tree (DAT) for information aggregation.

(i) In MAAN, resource sites are registered with a set of attribute value pairs and can be searched through multi-attribute based range queries. To address a range query, MAAN routes the query to the first resource site within its domain range. Then the query is forwarded to other resource sites consecutively until it reaches the last site.

(ii) In DAT, all resource sites use a well-adjusted routing algorithm to build a balanced DAT tree in the direction of the root site. In DAT, all nodes aggregate towards the global information.

Merits:

(i) Message overhead can be reduced.

(ii) Large resource network can be managed using hierarchical structure.

Demerits:

(i) Routing problems occurred.

(ii) Network traffic is increased.

(iii) Non-Standard protocols.

(f) Grid Information Services for Distributed Resource Sharing [12]:

Description:

Grid Information Services are providing fundamental mechanisms for resource discovery and monitoring using Information provider. Here Information provider expresses two protocols.

(i) GRIP (GRid Information Protocol) is used to access the information or characteristics about resource entities.

(ii) GRRP (GRid Resource Protocol) is used to perform notifying messages for aggregate directory services to find the availability of resource information.

Merits:

(i) These protocols are highly scalable for designing the Virtual Organizations (VO).

(ii) This approach defines some access control policies to provide security to the resources.

(iii) The advertisement and discovery of networked resources can be easily done with SLP (Service Location Protocol).

Demerits:

(i) Implementation of directories is a tedious process & low scalability.

(g) Information Aggregation through Omnipresent scheme [13]:

Description:

This approach defines omnipresent information aggregation scheme. Omnipresent means that every node in the grid network has a view of the aggregated information in a system and therefore called n-aggregation (n stands for n participants). Thus the every node is monitoring the networked resources.

Merits:

(i) All nodes are able to update information about the whole resource network.

Demerits:

(i) The aggregated messages might be loss due to poor communication channels.

(ii) Failures of nodes might be happened.

(h) Practical Private Information Aggregation [14]:

Description:

This model represents protocols for confidentially computing a large collection of aggregation functions based on addition, disjunction, and max/min. It encapsulates the resource information and uses some mathematical structure to ensure the protection of information. This paper defines secure protocols for aggregating the resource information.

Merits:

(i) More secure.

(ii) Highly scalable.

Demerits:

(i) Implementation is very complicated.

(ii) Data processing takes more time.

(i) Data Aggregation and Analysis: A Grid based approach for Medicine and Biology [15]:

Description:

This paper presents a Grid based approach to aggregate the medical related and biological information. Here mainly two components are used.

(i) Data Aggregator is able to aggregate the medical information deriving from heterogeneous resources. (E.g. biosensors, actuators)

(ii) Data Analyser deploys the information collected from the Aggregator component and takings with a set of simulations to produce the personalized healthcare and medicine scheme.

(j) Context Information Aggregation Scheme [16]:

Description:

This model uses Quality of Context parameters to resolve the conflicts in context information. Here the context aggregation system notices conflict and duplicate information which are kept in GIS (Grid Information System). If any conflict arises, the system removes that critical information using some quality aware policies. All context resource information is received at a monitor and aggregates the information from lower level schedulers before forwarding it to the higher level schedulers. The context information model suggests two algorithms.

Algorithm 1 detects duplicates and conflict objects in its Information database.

Algorithm 2 resolves conflict data.

Merits:

(i) High QoS (Quality of Service) is achieved.

(ii) This method removes repeatable information. So it conserves the storage space.

Demerits:

(i) Finding threshold values will be a difficult one.

(ii) Complicated architecture and it needs high-end infrastructure support. Therefore managing of resources will be a difficult one.

III. Comparative analysis of various resource information aggregation schemes in grid computing

	PARAMETERS/ PAPERS	OBJECTIVE	MERITS	DEMERITS
1	Domain based resource information aggregation	Using this approach, the grid network is partitioned into domains and each domain monitor computes its aggregated information.	(i) Resource monitoring is easy. (ii) Minimizing clumsy data.	(i) Efficient Data Structure is needed.
2	Information aggregation through BestBrokerRank policy	The central scheduler selects the appropriate domain which has best rank using aggregated information.	(i) High accuracy. (ii) Bandwidth saved. (iii) Highly scalable.	(i) Execution time is longer. (ii) Resource information size is larger.
3	Data Consolidation and Information Aggregation	Data consolidation method introduces to transfer group of data among the resource sites and data replicas.	(i) Grouping the domains can be easy. (ii) Reduces network load.	(i) Effective routing algorithm is needed. (ii) Every node needs more space queues.
4	Grid Information Services for Distributed Resource Sharing	Grid Information Services are providing fundamental mechanisms for resource discovery and monitoring using Information provider.	(i) Resource discovery can be easy. (ii) Effective access control.	(i) Implementation of directories is a tedious process. (ii) Low scalability.
5	Information Aggregation through Omnipresent scheme	In this model, every resource node in grid network monitors other resource nodes and computes its aggregated information.	(i) Information updating is easy.	(i) Node failures can be happened. (ii) Poor communication channels.
6	Aggregating Service with Failure Detector	This method introduces fault detection architecture called HYDRA to monitor the process failures in distributed systems.	(i) Network failures can be detected easily. (ii) High Availability.	(i) Complicated architecture. (ii) Low protocol security.
7	Practical Private Information Aggregation	This scheme defines secure protocols for aggregating the resource information.	(i) More secure. (ii) Highly scalable.	(i) Implementation is very complicated. (ii) Data processing takes more time.
8	Distributed Indexing based aggregation scheme	This method proposes a Multi-Attribute Addressable Network (MAAN) for resource indexing and a Distributed Aggregation Tree (DAT) for information aggregation.	(i) Message overhead can be reduced. (ii) Managing of resources is easy.	(i) Routing problems occurred. (ii) Network traffic is increased. (iii) Non-Standard protocols.
9	Context Information Aggregation Scheme	This aggregation scheme uses Quality of Context parameters to resolve the conflicts in context information which are kept in GIS.	(i) High Quality of Service is achieved. (ii) It conserves the storage space.	(i) Needs Complicated architecture. (ii)Resource management is difficult.

Table – I: Comparative analysis of various resource information aggregation schemes in grid computing.

IV. Conclusion

In this paper, various resource information aggregation schemes in grid computing have been surveyed. Even though discussed methods in Grids that consider aggregation as an available mechanism, they typically consider specific policies and scenarios, and they do not address through appropriate metrics the effect information aggregation has on the quality of the scheduling decisions made. This work handles information aggregation in grids as a distinct and important problem, trying to identify the main issues, limitations, dependencies and side-effects related to the aggregation process. The future work will be based on the above findings to develop a better optimized method to aggregate the resource related information for resource discovery.

References

[1] http://computer.howstuffworks.com/grid-computing1.htm
[2]http://www7.hp.com/enterprise/w1/en/technologies/grid-computing-overview.html
[3] Ian Foster, Carl Kesselman, Steven Tucke "The anatomy of the Grid".
[4] Klaus Krauter, Rajkumar Buyya, Muthucumaru Maheswaran "A taxonomy and survey of grid resource management systems for distributed computing".
[5] Serafeim Zanikolas, Rizos Sakellariou "A taxonomy of grid monitoring systems".
[6] P.Kokkinos, E.A.Varvarigos " Scheduling efficiency of resource information aggregation in grid networks".
[7] P.Kokkinos, E.A.Varvarigos "Resource Information Aggregation in Hierarchical Grid Networks".
[8] Ivan Rodero, Francesc Guim, Julita Corbalan, Liana Fong, "Grid broker selection strategies using aggregated resource information".
[9] P.Kokkinos, E.A.Varvarigos "Data Consolidation and Information Aggregation in Grid Networks".
[10] Yunhee Kang "A Process Status Aggregating Service with Failure Detector".
[11] Min Kai "Distributed Indexing and Aggregation techniques for Peer-to-Peer and Grid Computing".
[12] Karl Czajkowski, Steven Fitzgerald, Ian Foster, Carl Kesselman "Grid Information Services for distributed resource sharing".
[13] Mark Schneider "Information Aggregation for load balancing in a Distributed System of Grid Services".
[14] Kreitz, Mads Dam, Duglas Wikstorm"Practical Private Information aggregation in large networks".
[15] Dimosthenis Kyriazis, Konstantinos Tserpes, George Kousiouris, Andreas Menychtas, Gregory Katsaros, Theodora Varvarigou "Data Aggregation and Analysis: A Grid based approach for Medicine and Biology".
[16]. Atif Manzoor, Hong-Linh Truong, Schahram Dustdar "Quality Aware Context Information Aggregation System for Pervasive Environments".

Authors

R. Venkatesan,
Assistant Professor SG,
Department of Information Technology,
School of Computer Science and Technology,
Karunya University,
India.

M. Manicka Raja,
PG Scholar,
Department of Information Technology,
School of Computer Science and Technology,
Karunya University,
India.