International Journal of Grid and Distributed Computing
Vol. 4, No. 3, September, 2011

A Taxonomy of Grid Resource Selection Mechanisms

Adil Yousif, Abdul Hanan Abdullah,
Muhammad Shafie Abd Latiff and Mohammed Bakri Bashir

Faculty of Computer Science & Information System
Universiti Teknologi Malaysia UTM, Malaysia

Abstract

Resources in grid systems are heterogeneous, geographically distributed, belong to different administrative domains and apply different management policies. The roles of resource selection mechanisms are to identify, select and allocate the most suitable resources for a given set of tasks. This paper presents a taxonomy that facilitates identifying and classifying the mechanisms used in the implementation of grid resource selection process, as well as describing the most significant features of grid resource selection mechanisms. The benefit of this taxonomy is to highlight the main aspects of the selection mechanisms, which can help researchers and developers of grid resource management systems.

Keywords: *Taxonomy, grid computing, resource selection and resource allocation.*

1. Introduction

Grid computing emerged in the middle of 1990s as a wide-scale distributed system to offer dynamic coordinated resources sharing and high performance computing[1]. Grid technologies have evolved over the past two centuries from primary metacomputing into open grid service architecture (OGSA) using service oriented architecture (SOA) Concepts.

Resources in grid system are heterogeneous, geographically distributed, belong to different administrative domains and apply different management policies[1]. Furthermore grid management systems do not have full control over the resources that belong to the grid. The term resource is defined in grid context to denote any capability that may be shared and exploited in a networked environment. The resources and services may differ in form of functionalities that they offer to the user, but both are similar in the way that they provide those functionalities to the users [1-3]; hence we can use the term resource more generally to refer to all types of ancient resources, as well as services.

The taxonomy presented in this paper helps in identifying and classifying the mechanisms used in the implementation of grid resource selection process, as well as describing the most significant features of grid resource selection mechanisms. In this taxonomy we aim to highlight the main aspects of the selection mechanisms, which can benefit researchers and developers of grid resource management systems. The rest of the paper is organized as follows. Section 2 illustrates the related works, followed, in Section 3, by the description of phases of resource allocation process in grid computing. Section 4 presents the taxonomy that categorizes the resource selection mechanisms. We conclude the paper in Section 5.

2. Related Works

Considering taxonomies in grid resource management process, Krauter et al [2]presented a taxonomy of grid resource management systems; this taxonomy mainly focused on categorizing grid systems and the whole process of resource management in grid. However, this study does not focus on resource selection in details. In [3] taxonomy the authors described the basic features of resource brokers and grid middlewares. Also a taxonomy of grid application is presented in[4]. Nassif et al In [5]illustrates a taxonomy and new resource selection system; however this taxonomy and system are based on decision theory, case-based reasoning, and fine-grain policies. Our taxonomy in this paper is focusing mainly in the resource selection mechanisms and aims to help in identifying and classifying the mechanisms used in the implementation of grid resource selection process; as well as describing the most significant features of grid resource selection mechanisms.

3. Phases of Resource Allocation Process

The resource Allocation process in grid systems as shown in figure 1 consists of three main phases: resources discovery, resource selection and resource usage. Following is a brief description of the resource allocation phases.

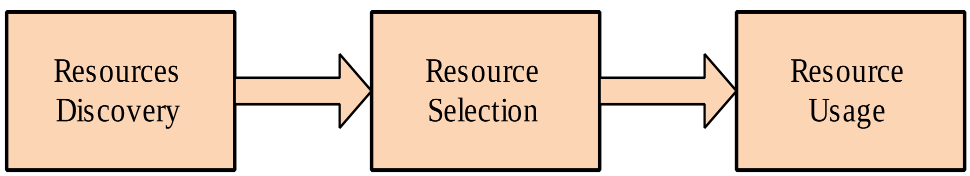

Figure 1. Resource Allocation Phases

3.1 Resource Discovery

Resource Discovery is the process of finding the appropriate resources that match the users job requirements [6]. One of the most important requirement for grid systems is an efficient resource discovery technique, which helps managing the resources and scheduling the users jobs. The activities of the resource discovery in grid systems involve resource registration in directory services and searching for the suitable resource types that match the application needs. The discovery process returns a set of candidate resources that can be used to execute the user application. A range of solutions has been proposed for resource discovery, including the hierarchical, peer-to-peer and centralized approaches.

3.2 Resource Selection

The second phase of resource allocation process is selecting a number of resources that have been discovered in phase one resources to perform the users applications[7]. In some cases, resource selection phase is interleaved with the discovery phase to collect more information on the resources that have met the requirements[8]. As a result in some resource management approaches, the discovery and selection phases are twisted together and recognized in one single element. As such, it is so difficult sometimes to find out differences between resource discovery and selection phases.

3.3 Resource Usage

This phase focuses in running the users applications on the selected resources and monitoring the execution. After Service Level Agreement (SLA) is initiated between the grid client and the resource provider, tasks can be implemented via the allocated resource. The resource provider may reserve some resources to achieve QoS in the agreement; also in some cases data is transferred from the grid client to the resource provider in order to process it with GridFTP. Furthermore some output files may move from the resource provider to the grid client if any. In this phase also the user can monitor the tasks during execution and when the task is ended, the grid client is informed by the provider[9].

4. Taxonomy of Resource Selection Mechanisms

Grid resource selection mechanisms can be categorized into several classes based on a number of features and characteristics. We present a taxonomy for grid resource selection mechanisms which covers mainly the common and important aspects and issues of grid resource selection mechanisms. The taxonomy presented in this paper describes the approaches used in selection decisions, the orientation and the objective in the selection process. In subsection a brief description of the selection models and methods is highlighted; this followed in subsection 4.5 by the organization of the selection and selection heuristics is presented in subsection 4.6. Classes of selection optimization methods are illustrated in subsection 4.7.

4.1 Selection Decisions

The decisions of selecting certain resources to perform client tasks are classified into two types as shown in Figure 2. Manual Selection is used when the grid clients desire to have full control over the selection process [9]. Often manual decision is used either if there are a small number of resources available in the network such as in a home Grid[10], or when the requested services are very rare and difficult to be described by the Resource description Languages RDL. Furthermore, Manual selection can be used when the resources or requested services are very rare; for instance, if a user requested a specific device for the detection of cancer by a physician. This device has specializations that are difficult to be described by resource description languages RDL. In this case, it would be desirable to give the physicians the list of available resources and let him choose the appropriate resources[9].

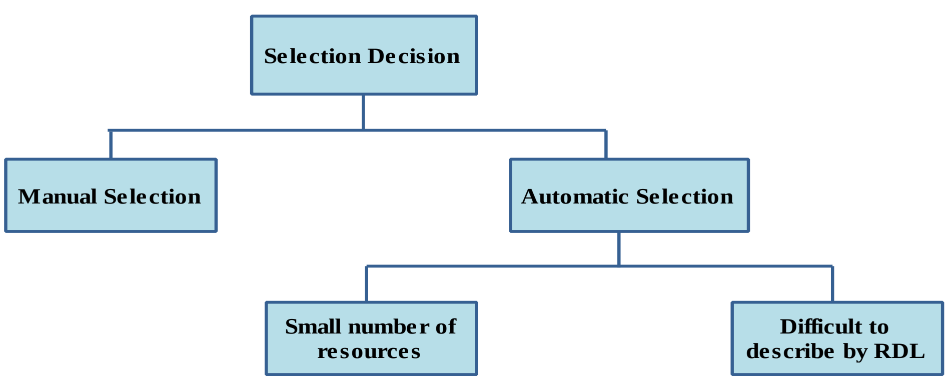

Figure 2. Selection Decisions

On the other hand, automated selection is suitable for situations where the grid networks have a huge number of resources and there are some RDLs that can describe those resources[11]. To apply the concept of Scalability, resource managers have to support the concept of automation. Since the grid systems consist of a large scale distributed resources, the automation of resource selection decisions had become a very critical feature.

4.2 Selection Orientation

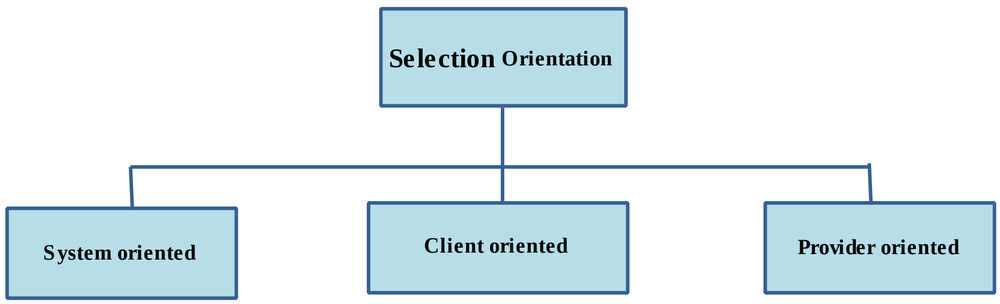

Figure 3. Selection Orientation

The grid resource selection methods are classified, as shown in Fig 3, into system oriented, grid client oriented and provider oriented[3]. In System oriented methods, such as Condor-G, the selection methods aim to optimize the overall system performance, average response times(ART), fairness and utilization[2, 12-13]. Whereas in the client oriented methods, such as methods implemented in EMPEROR[14], the selection algorithms focus on meeting the needs of individual client by minimizing the completion time for all tasks submitted by that client, regardless of the overall system performance and utilization[12, 15-16]. The provider oriented approach allows the resource providers to utilize and employ their resources effectively. However, this approach does not agree on any commitments for the status of their resources to any grid client[17].

4.3. Selection Models

In grid Resource selection mechanisms, there are two main types of resource selection models as described in figure 4.

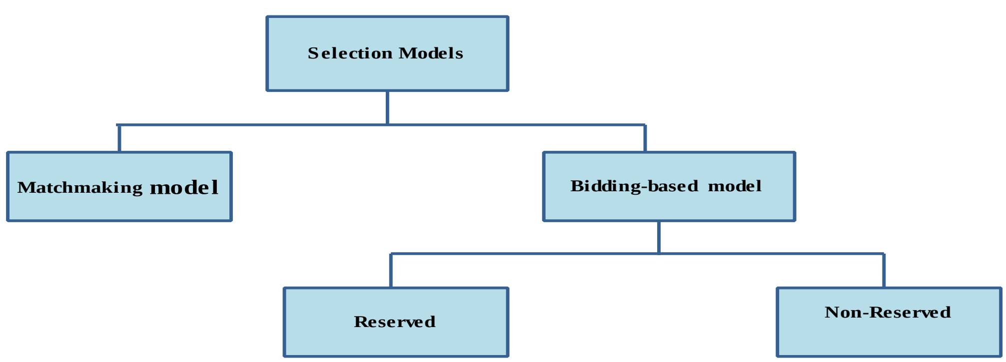

Figure 4. Selection Models

The functions of resource matchmaker (as shown in figure 5) are registering all resources status announced by resource providers and running matching algorithms. The rise in number of resources and the regularity of tasks demands lead to matchmaker overload problems. Furthermore, the matchmaker information often is out of dates. This is because the grid resource's status is changing frequently and the matchmaker does not learn about the resources status until the resources advertise their new status to the matchmaker [18-20].The matchmaker can be implemented by a job manager[21] or using a centralized resource broker as discussed in[12]. Moreover, the matchmaker, like other centralized systems, represents performance bottleneck and single point of failure.

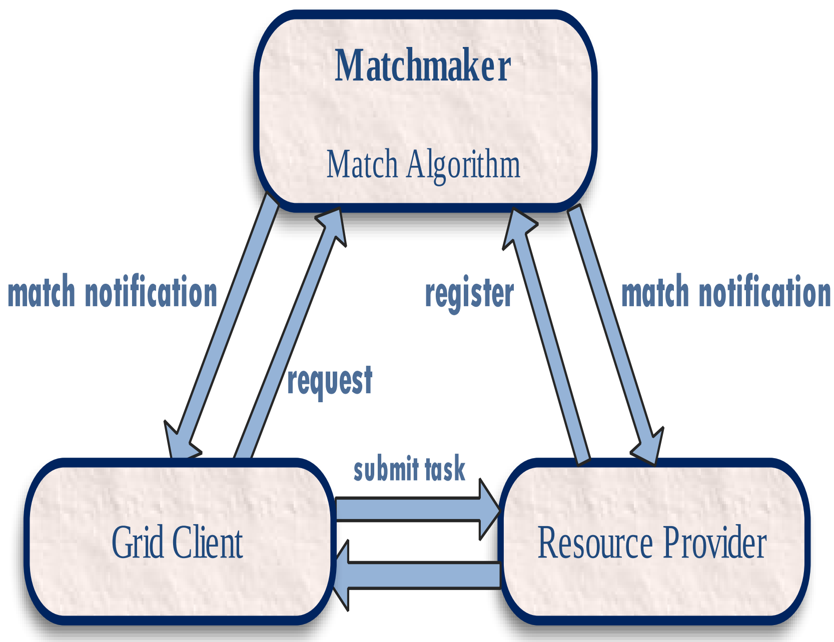

Figure 5. Resource Matchmaker

Bidding-based model emerged to address the expired information, and the performance problems arose when using the matchmaking model [18-21]. As described in figure 6, the bidding process starts when a grid client sends call-for-proposal (CFP) requests to all available resource providers. According to their characteristics and status of their resources, resource providers determine if they can join the bidding process or not. If a resource provider participates in a bidding process, it sends a bid that expresses the status of its resources to the grid client. The grid client evaluates the bids received from providers, orders them and selects the resources that offer the best bids [17-18, 22-23]. Bidding based model suffers from lack of global system information, and hence the client brokers make the selection decision based on partial system global state information. Furthermore, since bidding based models do not maintain centralized resource management method. The system state may become imbalance if many grid clients greedily select the same resource providers that have the highest computational power [18].

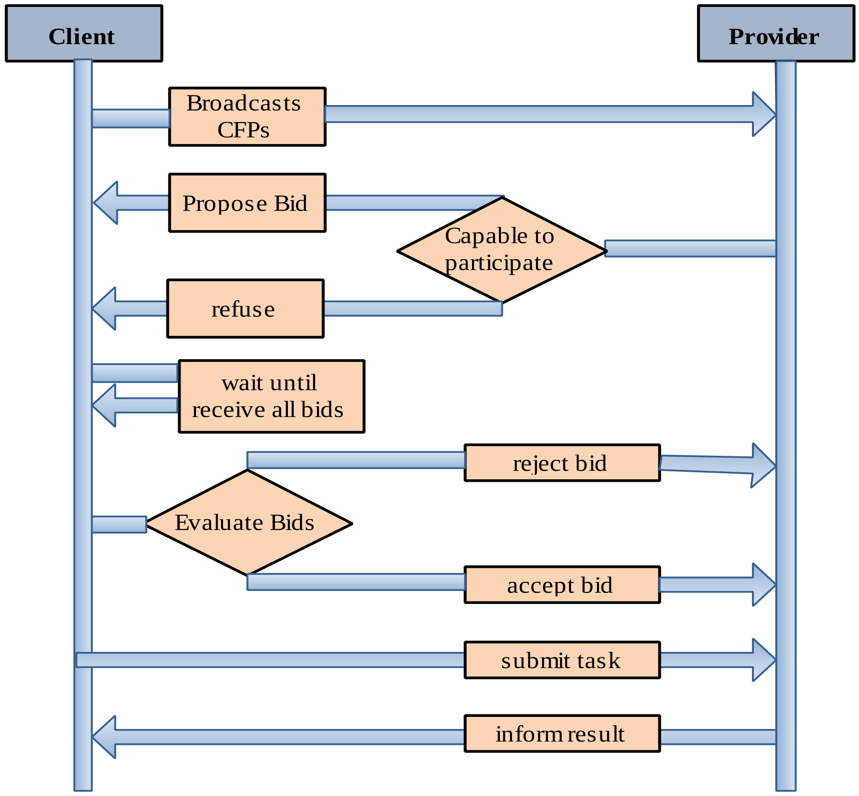

Figure 6. Bidding process

4.4 Securing Methods

Based on the commitments of resources for tasks and the usability of resources we can categories two types of the bidding resource securing or reserving model, reserved resource model and non-reserved resource model (figure 4). In the reserved resource model when a resource provider decides to join a bidding process, it reserves the resource for that process to guarantee resource status in the future[21]. However, if the grid client later refuses the resource, that resource is squandered. Under these circumstances, other grid users possibly are ready to agree to the resource prior to the original user refuses it. Therefore, the likelihood for the resource provider to lose the chance for the participation in those other processes and support those grid clients is high[24-25].

On the contrary, in the non-reserved model, the resource providers do not reserve the resources for any bidding process. This approach permits the resource providers to utilize and employ their resources effectively. However, it does not agree commitments for resources' status to any bid. In this case, a problem arises if more than one grid client send CFP for the same resources in the same resource provider simultaneously. The resource provider will participate in all bidding processes but will not guarantee the resource status for any client and hence the job execution time possibly will not be as estimated [18, 21, 25-26]. Furthermore, the non reserved model in most cases does not suit the economic grid environments, in which the grid client is obliged to pay for resource usage. For instance, if a grid client paid for specific resources, and the completion time is a critical factor to the user, then the resource reservation is essential to guarantee that the selected resources are committed to that user[13, 26].

Reserved resource model wastes the providers' resources; this is due to, more than one resource providers reserve their resources for a single bidding process and at the end, the grid client selects only one resource for task execution. And hence other resources miss the chance to participate in other bidding processes and serving other clients before they are rejected by the original grid client. On the other hand, non-reserved resource model lead to unexpected completion time for tasks. This happens because the resource provider may participate in more than one bidding by the same resources simultaneously. This makes the grid clients compete for the resources, and this competition may cause unexpected completion time for the submitted tasks.

4.5 Organization

According to [2-3] the organization of resource selection mechanisms is either centralized or distributed (Figure 7). In centralized resource Selection mechanisms, there is a broker or server that handles and deals with all tasks submitted to the grid system[12]. One of the advantages of the centralized resource selection mechanism, it has knowledge and control over all grid resources and tasks submitted to the grid system. Therefore, centralized selection can provide a good management for the submitted tasks. However, centralized mechanisms may become a performance bottleneck. Furthermore, centralized resource selection mechanisms, like other centralized systems, usually represents a single point of failure [27].

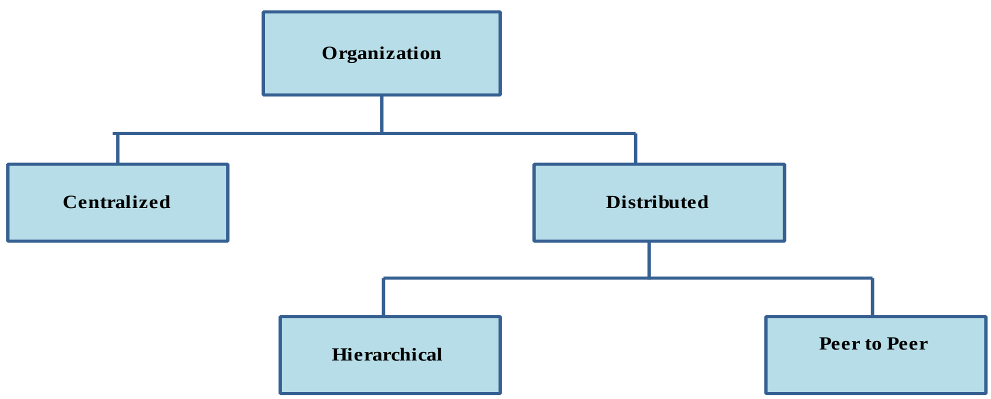

Figure 7. Selection Organization

In distributed selection mechanisms the selection process is distributed over all grid resources and there is no one server that controls all the grid network resources. Distributed selection mechanisms provide scalable grid systems and do not represent single point of failure. However, distributed organizations suffer from the lack of knowledge about the global state of the system, and do not have full control over the grid resources [7, 12]. Distributed mechanisms are either, peer-to-peer or hierarchical. In peer-to-peer(as shown if figure 8), all resources in organizations join the grid have to register in one or more index servers[12]. The index server itself can register in other index servers. All users in the grid system have their own brokers, and they can use them to access the available resources[2]. For each resource broker to discover which resources are available it communicates with one or more index servers. In the hierarchical organization, the hierarchy consists of several clusters; each cluster has a number of nodes and a root node, called the cluster root. The cluster root responsibilities are to control the brokering and selection of the corresponded cluster nodes and handles tasks submitted to the cluster[28-29]. To construct a higher level in the hierarchy, a number of clusters are collected and controlled by a root node[2].In this organization if any resource participates or leaves the virtual organization its status is gathered and handled by the higher level cluster root. The key feature of hierarchical schemes is to perform local resource selection and brokering first (within the cluster). If the selection or brokering algorithm fails to find matched resources within the cluster, then it searches in other levels of the hierarchy, so the aim of hierarchical organization is to reduce the amount of traffic and data exchange between clusters[28].

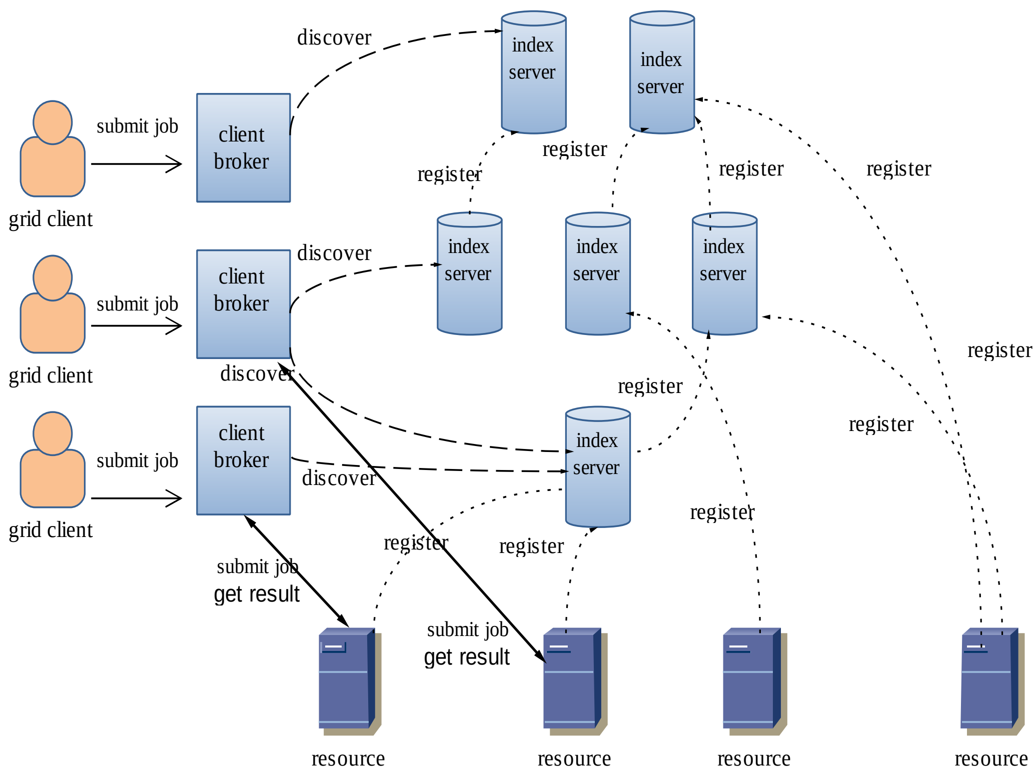

Figure 7. Distributed Selection

4.6 Selection Heuristics

In the process of resource selection in heterogeneous environment, it is difficult to find rules that determine the right selection decisions. However researchers use heuristics to help obtaining better selection decisions. Heuristics are approaches that help make right decisions; but they do not always produce the correct selection decisions[30]. There are two different types of heuristics, as shown in Fig 8, that are commonly used in the process of resource selection; deterministic and probabilistic or stochastic[18]. In deterministic heuristics, all resources states are uniquely determined by parameters in the heuristics model. However, in probabilistic heuristics variables states are not determined uniquely, but they determined using probability distributions. Hence deterministic heuristics achieve the same result for a given initial conditions which is not necessary occur when using probabilistic heuristics[18].

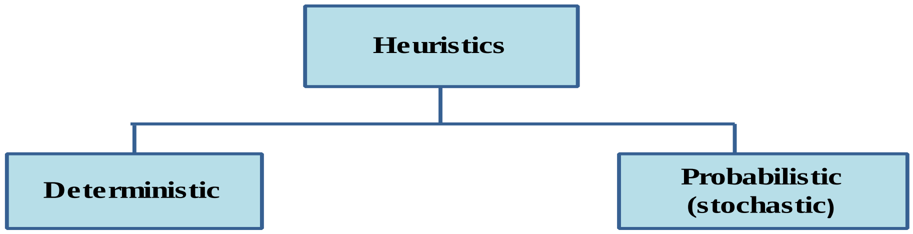

Figure 8. Selection Heuristics

4.7 Mapping Methods

The process of mapping jobs to resources and figuring out the execution order of the jobs allocated to each resource is classified into static and dynamic, figure 9. Dynamic techniques do the mapping immediately when jobs arrive. Whereas in static techniques, the total set of jobs is identified a priori and the mapping process is done previous to the execution of any of the jobs [31]. There are two types of dynamic mapping: on line mode and batch mode. In online mode the task is assigned immediately when it arrives the mapper; whereas in the batch mod tasks are gathered into a set, and this set is analyzed for resources s assigning before scheduling [31].

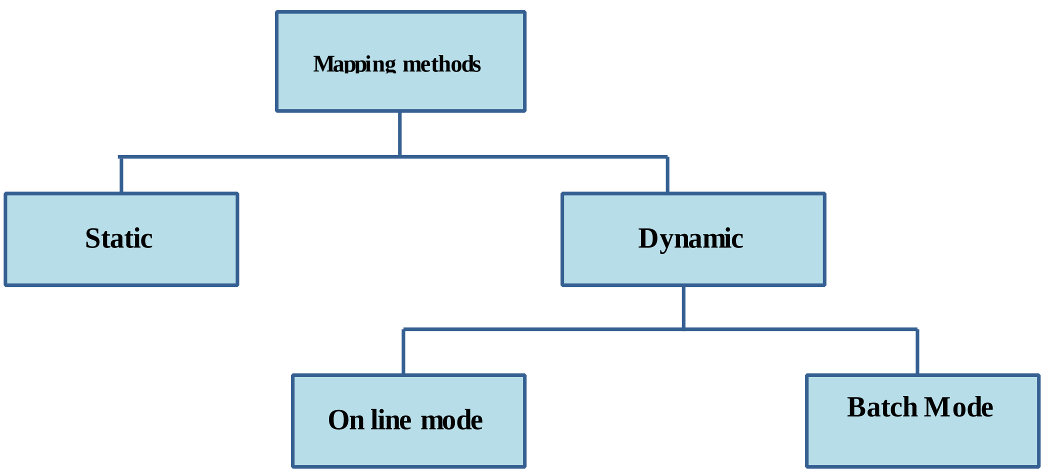

Figure 9. Mapping Methods

5. Conclusion

Having a consistent taxonomy enables researchers and developers to identify obviously in which categories their work falls in. This paper presented a taxonomy for grid resource selection mechanisms. The taxonomy has mainly focused on selection decisions, orientation, models, organization and mapping methods of resource selection mechanisms. Our Taxonomy describes, in a simple way, the basic features of resource selection mechanisms, which can help researchers and developers of grid resource management systems to enhance the grid resource allocation process.

References

[1] Foster, I. and C. Kesselman, *The grid: blueprint for a new computing infrastructure*. 2004: Morgan Kaufmann.

[2] Krauter, K., R. Buyya, and M. Maheswaran, *A taxonomy and survey of grid resource management systems for distributed computing.* Software: Practice and Experience, 2002. **32**(2): p. 135-164.

[3] Kertész, A. and P. Kacsuk, *A taxonomy of grid resource brokers.* Distributed and Parallel Systems, 2007: p. 201-210.

[4] Suciu, A. and R. Potolea, *A taxonomy for grid applications.* 2008.

[5] Nassif, L.N., J.M. Nogueira, and F.V.V. De Andrade, *Resource selection in grid: A taxonomy and a new system based on decision theory, case-based reasoning, and fine-grain policies.* Concurrency Computation Practice and Experience, 2009. **21**(3): p. 337-355.

[6] Iamnitchi, A. and I. Foster, *On fully decentralized resource discovery in grid environments.* Grid Computing—GRID 2001, 2001: p. 51-62.

[7] Chang, R.S., C.F. Lin, and J.J. Chen, *Selecting the most fitting resource for task execution.* Future Generation Computer Systems, 2011. **27**(2): p. 227-231.

[8] Schopf, J., *Ten actions when grid scheduling.* International Series in Operations Research and Management Science, 2003: p. 15-24.

[9] Schnizler, B., *Resource Allocation in the Grid: A Market Engineering Approach*. 2007: Univ.-Verl. Karlsruhe.

[10] Meliksetian, D., et al., *Design and implementation of an enterprise grid.* IBM Systems Journal, 2010. **43**(4): p. 646-664.

[11] Clematis, A., et al., *Job-resource matchmaking on Grid through two-level benchmarking.* Future Generation Computer Systems, 2010. **26**(8): p. 1165-1179.

[12] Elmroth, E. and J. Tordsson, *Grid resource brokering algorithms enabling advance reservations and resource selection based on performance predictions.* Future Generation Computer Systems, 2008. **24**(6): p. 585-593.

[13] Haji, M.H., et al., *A SNAP-based community resource broker using a three-phase commit protocol: A performance study.* Computer Journal, 2005. **48**(3): p. 333-346.

[14] Adzigogov, L., J. Soldatos, and L. Polymenakos, *EMPEROR: An OGSA Grid meta-scheduler based on dynamic resource predictions.* Journal of Grid computing, 2005. **3**(1): p. 19-37.

[15] Elmroth, E. and J. Tordsson, *A standards-based Grid resource brokering service supporting advance reservations, coallocation, and cross-grid interoperability.* Concurrency Computation Practice and Experience, 2009. **21**(18): p. 2298-2335.

[16] Balakrishnan, P. and T.S. Somasundaram, *SLA enabled CARE resource broker.* Future Generation Computer Systems, 2011. **27**(3): p. 265-279.

[17] Yousif, A., A.H. Abdullah, and A.A. Ahmed, *A Bidding-based Grid Resource Selection Algorithm Using Single Reservation Mechanism.* International Journal of Computer Applications, 2011. **16**(4): p. 39--43.

[18] Wang, C.-M., et al., *Dynamic resource selection heuristics for a non-reserved bidding-based Grid environment.* Future Generation Computer Systems, 2010. **26**(2): p. 183-197.

[19] Khanli, L.M. and M. Analoui, *An approach to grid resource selection and fault management based on ECA rules.* Future Generation Computer Systems, 2008. **24**(4): p. 296-316.

[20] Yu, H., X. Bai, and D.C. Marinescu, *Workflow management and resource discovery for an intelligent grid.* Parallel Computing, 2005. **31**(7): p. 797-811.

[21] Hongbo, Z., et al. *A virtual-service-domain based bidding algorithm for resource discovery in computational grid*. in *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*. 2005.

[22] Kakarontzas, G. and I.K. Savvas. *Agent-Based Resource Discovery and Selection for Dynamic Grids*. in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE '06. 15th IEEE International Workshops on*. 2006.

[23] Kolano, P.Z. *Surfer: an extensible pull-based framework for resource selection and ranking*. in *Cluster Computing and the Grid, 2004. CCGrid 2004. IEEE International Symposium on*. 2004.

[24] Cheung, W.K., et al. *Dynamic Resource Selection For Service Composition in The Grid*. in *Web Intelligence, 2004. WI 2004. Proceedings. IEE E/WIC/ACM International Conference on*. 2004.

[25] Xueguang, C. and S. Haigang, *Further extensions of FIPA Contract Net Protocol: threshold plus DoA*, in *Proceedings of the 2004 ACM symposium on Applied computing*. 2004, ACM: Nicosia, Cyprus. p. 45-51.

[26] Haji, M., K. Djemame, and P. Dew. *Deployment and performance evaluation of a SNAP-based resource broker on the White Rose grid*. 2006: IEEE.

[27] Roy, S., M. Sarkar, and N. Mukherjee. *Implementation of a Resource Broker for Efficient Resource Management in Grid Environment*. in *Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on*. 2007.

[28] Malarvizhi, N. and R. Uthariaraj. *Hierarchical load balancing scheme for computational intensive jobs in Grid computing environment*. 2010: IEEE.

[29] Huedo, E., R. Montero, and I. Llorente, *A recursive architecture for hierarchical grid resource management.* Future Generation Computer Systems, 2009. **25**(4): p. 401-405.

[30] Resende, M.G.C., J.P. de Sousa, and A. Viana, *Metaheuristics: computer decision-making*. 2004: Springer Netherlands.

[31] Magoules, F.F., *Grid resource management : towards virtual and services compliant grid computing*, ed. C.H.C.n.a.a.s. computing. 2009: Boca Raton :CRC Press.

Authors

Adil Yousif Received the B.Sc. and M.Sc. from University of Khartoum, Sudan. He is a lecturer at Faculty of Computer Science and Information Technology, Kassala University. Currently he is a PhD candidate at Faculty of Computer Sciences & Information Systems, Universiti Teknologi Malaysia UTM. He is also member of PCRG research group. His research interests include computer network, distributed systems and grid computing.

Prof. Abdul Hanan Abdullah Received the B.Sc. and M.Sc from San Francisco, California and his PhD degree from Aston University in Birmingham, United Kingdom in 1995. He is now a Professor at the Faculty of Computer Science and Information systems, Univeristi Teknologi Malaysia UTM. Currently he is heading Pervasive Computing Research Group, a research group under K-Economy Research Alliances. His research interests include computer network, network security and grid computing.

Muhammad Shafie Abd Latiff received the B.S. 1986 and M.S. 1995 degrees in Computer Science (Universiti Teknologi Malaysia) and PhD in Modeling, Simulation and Animation in Virtual Environments (Bradford University, United Kingdom) in 2002. He is now a Associate Professor at Faculty of Computer Science and Information systems, Univeristi Teknologi Malaysia UTM.

Mohammed Bakri Bashir Received the B.Sc. and M.Sc. from SUST University, Sudan. He is a lecturer at University of Shandi. Currently he is a PhD candidate at Faculty of Computer Sciences & Information Systems, Universiti Teknologi Malaysia UTM. He is also member of PCRG research group. His research interests include computer network, distributed systems and grid computing.