



Research Article

A Novel Optimization Model Based on the Unification of Proximity and Semantic Similarity in Grid Computing

Abdul Khalique Shaikh¹, Saadat M. Alhashmi², Rajendran Parthiban³

¹Artificial Intelligence Laboratory MIMOS Berhad TPM , Bukit Jalil, Malaysia

²College of Engineering & Computer Science, Abu Dhabi University, Abu Dhabi UAE

³School of Engineering, Monash University, Sunway Campus, Bandar Sunway, Malaysia

Correspondence should be addressed to: Abdul Khalique Shaikh; shaikhned@yahoo.com

Received date: 5 September 2014 ; Accepted date: 22 February 2015; Published date: 25 November 2015

Academic Editor: Kalaivany Natarajan

Copyright © 2015. Abdul Khalique Shaikh, Saadat M. Alhashmi, Rajendran Parthiban. Distributed under Creative Commons CC-BY 4.0

Abstract

Resources in Grid computing are geographically distributed across the world through a wide area network under various virtual organizations. Due to the distributed nature of the Grid, the selection and allocation of the optimal resources from the available resource are challenging. However, the overall Grid performance depends on the selection of Grid resources for user jobs. A significant amount of effort has been made by proposing various resource discovery algorithms. Current Grid literature reveals that the semantic matching can provide more results compared to syntax matching on available resources, but the selection of poor resources for user jobs can affect the Grid performance. The reason for poor selection is because of the allocation of Grid resources based on First Come First Serve (FCFS) scheme, which reduces the utilization of a domain-based semantic ontology Grid system. To overcome the issue and enhance the Grid performance, we propose a novel optimization model based on Unification of Proximity and Semantic similarity in Grid Computing. The purpose of this optimization model is to get optimized resources for user jobs, so that Grid brokers could select optimum resources in terms of proximity with high semantic relevancy. The proposed model utilizes both semantic and proximity criteria and avoids the resources that are not suitable and faraway from the user locations. The model is designed using GridSim and FreePastry simulation & modeling toolkits. The experimental results have been compared with the (FCFS) allocation scheme that shows that the proposed optimization model statistically significantly outperforms the system with FCFS scheme.

Keywords: Grid; Semantic; Proximity, GridSim; FreePastry; Resource Allocation; Decentralized Resource Discovery

Cite this Article as: Abdul Khalique Shaikh, Saadat M. Alhashmi, Rajendran Parthiban (2015), " A Novel Optimization Model Based on the Unification of Proximity and Semantic Similarity in Grid Computing ", Journal of Software & Systems Development, Vol. 2015 (2015), Article ID 926190, DOI: 10.5171/2015.926190

Introduction

Grid computing is an extremely large and distributed system where resources join and leave frequently. Due to the above characteristics, the resource selection and allocation of resources for user jobs is exceptionally challenging. Moreover, Grid resources reside under distinct virtual organizations with their own rules and policies (R. Ranjan & R. Buyya, 2009). By using semantic features in Grid computing environments, the resource availability can be enhanced that helps in the allocation of resources in the Grid. Prior to submitting jobs to a Grid system, appropriate resources are selected to execute user jobs. However, resources are highly distributed in a Grid system and are dynamic in nature. Ian Foster states in (Iamnitchi, Foster, & Nurmi, 2003) that unlike global identification of resources in distributed systems, it is extremely difficult to define a global naming scheme for attribute based resource identification in a Grid computing environment. In the light of the above condition, it is highly probable that the same resources might be published with different names and it could be possible to miss some relevant resources in syntax-based techniques. Hence, the selection and allocation of Grid resources in Grid computing are particularly challenging. Due to the usage of fixed schema between users requirement and providers availability in Grid environments, the job rejection ratio is extremely high. To overcome the above limitations, the usage of semantic technology is being considered to reduce the job rejection ratio, because semantic matching helps to remove the tight coordination between Grid resource providers and Grid users. The overall effectiveness of the system depends on the level of coordination and cooperation among users, providers, resources and services (R Ranjan & R Buyya, 2009). In order to enhance job success rate, better coordination between users and providers are required that could be achieved by adding semantic features.

Currently, the selection and allocation of Grid resource in the existing semantic models is based on FCFS basis. In this scheduling

technique, when the queries match the relevant resources, the model neither considers node proximity into account nor the best degree of semantic relevant resources. However, both factors could be offered at the time of Gridlet scheduling. The reason for this is that, the broker picks the first available semantic matched resources based on FCFS scheduling. Hence, the matched resource can be far away from user nodes in terms of proximity. Also, it is possible that the first matched resources do not always provide the best semantic relevant resource. By considering these issues, we select the best resource for user jobs among available resources by presenting an optimization model which is based on the unification of proximity and semantic similarity matching. The model can overcome the issue in the existing decentralized semantic resource selection and allocation models such as (Li, 2010; Liangxiu & Berry, 2008; Pirrò, Talia, & Trunfio, 2012), where a domain-based ontology is utilized that may provide semantic relevant resources, but that could be different in terms of function that can lead to the rejection of jobs at run time. The model optimizes the selection of an appropriate resource for current job requirements based on node proximity and semantic similarity factors. We evaluate the model in a combination of Grid-based simulator GridSim and network overlay simulator FreePastry. The proposed optimization model demonstrates significant improvement by considering proximity and semantic similarity in the selection of appropriate resources. The model is based on Proximity and Semantic similarity matching values for the optimization of selection and allocation of user jobs in decentralized Grid environment. The remaining sections of the paper are organized as follows:

Section 2 briefly explains the state of the art related works in the existing selection and allocation models in Grid Computing. Section 3 describes an overview and query process mechanism of the proposed optimization model. Section 4 discusses the semantic mapping and semantic matching in the proposed optimized model. Finally, Section 5 concludes the paper with possible future work.

Related Works

This section explains an overview of existing state of the art related works in the Grid system for the selection and allocation of Grid resources.

A scalable DHT and ontology based Information Service (DIS) has been proposed by (Tao, Jin, Wu, & Shi, 2009) by using Chord protocol. The approach is similar to super peer concept with a slight difference. The DIS support DHT query and semantic based query. The service avoids traversing all nodes and parsing each service description document, speeding up the query process and improving the query precision. The proposed solution has been evaluated on China Grid environment. The model measures scalability and query response time. Authors claim that the model can improve the query precision, high throughput and speed up the information query and supports high scalability. The main reason for using Chord P2P overlay protocol is that it provides efficient lookup and routing functionality with fast distributed computations of hash function. However, the author of Chord routing protocol stated in (Stoica et al., 2003) that the Chord routing information is not much efficient if the number of nodes is extremely high.

The paper (Liangxiu & Berry, 2008) has introduced a semantic supported agent-based decentralized Grid resource discovery mechanism. This heuristic algorithm finds out neighbor resources and introduces the concepts of semantic similarity through a domain ontology using a decentralized approach. The experimental results show that the job success probability of the resource discovery increases with the decrease in semantic threshold values. The authors in (Liangxiu & Berry, 2008) claim that the algorithm has the flexibility to discover resources in an efficient and dynamic way. However, the paper uses a domain-based ontology and the experimental results show that the job success probability is extremely low under average job complexity and average semantic threshold values.

By using Pastry DHT protocol, a simple approach has been presented to build a distributed content based publish/subscribe system in the research paper (Tam, Azimi, & Jacobsen, 2004). The paper uses a similar kind of RDBMS schema in this approach that helps to discover topics from the content of subscriptions and publications. The approach increases the expressiveness of subscriptions compared to the topic-based system. Based on evaluation, authors claim that it supports scalability and it could be possible to achieve accurate and efficient matching. However, it does not fully support query semantics of a traditional content-based system. In addition, the fault tolerance has not been evaluated in subscription storage.

The research work (Li., 2010) has proposed a semantic approach that is called OntoSum for efficient resource information integration and services. The OntoSum provides an efficient semantic search for resource discovery by implementing ontology domain knowledge and Semantic Link Network (SLN). A RDV-based (Resource Distance Vector) semantic routing element has been used for finding Grid resources in which semantic information is isolated into small chunks. The authors claim that OntoSum supports complex semantic web data, and it dramatically outperforms existing shortcut and network scheme in terms of scalability. However, proximity is not being considered in the selection of resources, and the results of the OntoSum are based on artificially generated data with a domain-based ontology.

A recent survey paper (Qureshi et al., 2014) identified that the existing resource discovery mechanisms need to be improved in terms of the transparent utilization of resources so that tasks could be executed without any limitation. Also, the research revealed that the various types of job requirements change dynamically that can affect the Grid performance and need to be addressed at run time.

An Efficient Routing Grounded On Taxonomy (ERGOT) (Pirrò et al., 2012) has been presented where authors employ an extended version of Distribution Hash Table (DHT) Chord protocol

to publish various services description using ontology concepts, and utilize Semantic Overlay Network (SON) for clustering of nodes to overcome the limitation of syntax based DHT search. The authors in (Pirrò et al., 2012) state that the DHT is limited to exact search and does not support semantic queries. However, it provides better scalability, whereas ERGOT enables semantic driven queries but is less scalable. Based on the simulation results, authors claim that the system enhances the efficiency of the searching mechanism in terms of accuracy and communication overheads. However, the paper focuses on a service discovery in the non-Grid context and uses WordNet generic domain-based ontology that can negatively affects the overall performance of the system.

A recent paper (Somasundaram, Govindarajan, Kiruthika, & Buyya, 2014) proposed a semantic-

enabled CARE Resource Broker (SeSRB) that provides a common framework to describe grid and cloud resources, and to discover them in an intelligent manner. Authors simulate the real data applications with semantic Grid environment. The results of the experiment show that the jobs submitted to the resource broker, job rejection rate is reduced while job success and scheduling rates are increased. However, existing research reveals that centralized and hierarchical resource discovery models can perform poorly for large-size Grids, because of various limitations.

Different from the above approaches, we propose and design an optimization model that considers both proximity and semantic similarity matching in the ontology-based decentralized resource discovery model for Grid computing.

Table 1 shows how the proposed model is different from the existing latest work in terms of the various functionalities. See Table 1.

Table 1 : Comparison of the optimized proposed model with OntoSum, ERGOT and SeSRB

Model Functionality/Features	OntoSum 2010	ERGOT 2012	SeSRB 2014	(The proposed optimization model 2015)
Model Structure	Decentralized	Decentralized	Hierarchical	Decentralized
Semantic	Yes	Yes	Yes	Yes
Proximity	No	No	No	Yes
Ontology data	Artificial	WordNet Generic domain-based ontology	Use cases of Real Applications	Realistic Grid resource ontology concepts
Grid-based Simulation Toolkit	No	No	Yes	Yes
P2P routing protocol	Resource Distance Vector Algorithm	DHT(Chord)+SON with domain-based ontology	N/A	Pastry+Semantic with sub-domain-ontology-based structure
Scalability	High	High	Medium	High
Matching Criteria	Based on Semantic Relevancy with FCFS	Based on Semantic Relevancy with FCFS	Based on Domain specific semantic matching and close matching with less	Based on unification of Proximity and Semantic similarity

			number of processor	
Performance Analysis	Recall, Overhead composition, Average length Path	Communication Overheads, Recall	Job Success rate , Job Rejection rate, Scheduling success rate, execution time and throughput	Proximity, Recall, Semantic Similarity, Latency

Table 1 shows how the proposed model is different from the existing work such as OntoSum (Li, 2010) ERGOT (Pirrò et al., 2012) and SeSRB (Somasundaram et al., 2014). OntoSum and SeSRB use semantic features, but no proximity criteria, whereas, our proposed optimization model considers both proximity and semantic features in the matching and selection process. As far as ERGOT is concerned, it is decentralized and supports scalability. However, the ERGOT does not consider proximity, and it uses a domain-based ontology.

The Optimization Model

An overview

The proposed optimization model consists of two stages. It identifies the all-possible matched nodes that can fulfil the user requests in the first stage and in the second stage, selects the best optimal resource node based on proximity and high semantic similarity matching values for the selection and allocation of resources for Grid user. The general overview of the model can be seen in

Figure 1

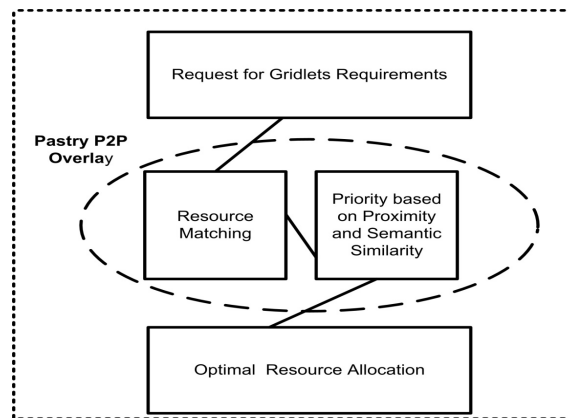


Figure 1:General Overview of Optimization Model

Figure 1 illustrates a general overview of the proposed optimization model that presents the unification of proximity and semantic similarity matching values based on a ranking method. By doing so, all the possible matched nodes are

collected for Gridlets based on the semantic similarity values and then the proximity between nodes is computed. The proximity distance is measured through scalar proximity metric in the Pastry P2P overlay, which is based on IP routing hops and the time required to

execute a query. After normalizing the values, the proposed model selects the optimized resource for user Gridlets. The main advantages of this unification are to reduce latency and improve the applications' performance. That can

also help in minimizing the network load. The process of the selection of an optimal resource in the proposed model can be seen in Figure2

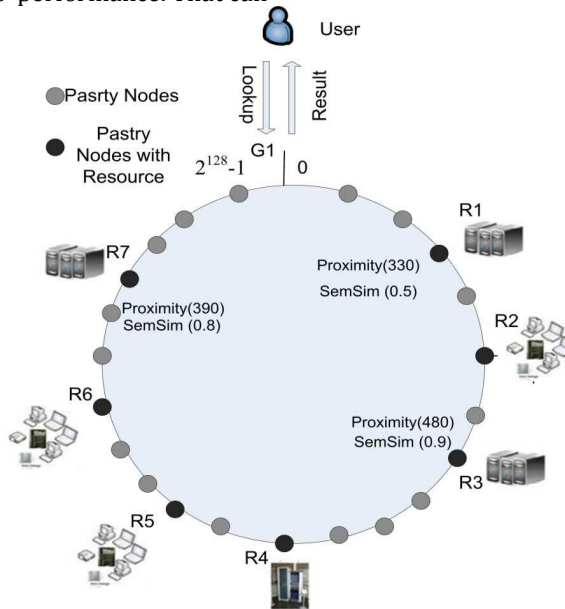


Figure2: Example of Selection of Optimal Resource

In Figure2, the Pastry ring contains seven resources where a user sends a query for Gridlet G1 requirements. It is assumed that R1, R3 and R7 are semantically matched resources that can fulfil the above Gridlet requirements. However,

before submitting the Gridlets, the model ranks the Grid resources based on the combination of the proximity criteria and the semantic similarity values and picks the best optimal resource as shown in the following table:

Table 2: Calculation of the Best Resource based on Proximity and Semantic Similarity

Matched Resource	Proximity (nearness in time)	Normalized Proximity	Semantic Similarity	Normalized Semantic Similarity	Ratio between Proximity and semantic similarity	Optimal value
R1	210	$210/340=0.61$	0.40	$0.40/0.80=0.5$	$0.61/0.5=1.22$	
R3	340	$340/340=1.0$	0.80	$0.80/0.80=1.0$	$1.00/1.00=1.00$	
R7	280	$280/340=0.82$	0.70	$0.70/0.80=0.87$	$0.82/0.87=0.94$	0.94 ✓

Table 2 shows the calculation procedure, how the proposed model picks the best resource among all matched resources based on proximity criteria and semantic similarities values. In the above example, if a broker allocates R1 for Gridlet that is closest to user's node, the semantic similarity is the lowest i.e. 0.40 and if resource R3 is allocated for current Gridlet then the user can get the advantage of the best semantic similarity but the resource is the farthest. In the light of the above situation, the model ranks all the possible matched resources and allocates resource R7 that is optimal in terms of both proximity and semantic similarity values where Gridlet G1 can be submitted. In other words, the optimal resource R7 can be used among the above three resources for G1 Gridlet that can lead to the enhancement of applications' performance in Grid Computing.

The proposed optimization model takes three inputs in the form of list (1) matched nodes that can fulfil the requirement of current Gridlet, (2) semantic similarity matching values of each match node and (3) proximity list of each matched node. The output is the best node based on the combination of both proximity and semantic similarity. First, the algorithm sorts the proximity list and gets the highest values. Then, it runs the loop and normalizes the values of the proximity list. Subsequently, it gets the highest value from the semantic similarity list and normalizes the value. After normalizing the values of semantic similarity matching and proximity elements, it measures the ratio between normalized values. After that, it sorts the ratio value and gets the original index respectively. Finally, it picks the best match resource based on the combination of the lowest proximity and the highest semantic similarity matching values and returns the best node resource to the user for running Gridlet.

Query Process

The proposed optimization model that is based on a combination of proximity and semantic similarity matching values has been designed to get the optimal resource for the Gridlets. In this sub-section, the process of resource query is

explained. First, a Pastry-based decentralized network has been established with various numbers of nodes using a FreePastry simulation toolkit. Then, the providers publish various specifications of Grid resources under these nodes. For simplicity, only one resource is assigned per node that contains multiple machines and multiple processors. Multiple Gridlets are created with different requirements using GridSim. After the network is stabilized, the Gridlets are sent to the Pastry network to identify the semantic relevant resources. The Pastry routing mechanism is used to route the query across the network that takes the semantic matched resource if an exact match is unavailable. The query collects the all-possible matched nodes and selects the best one based on a combination of the proximity and the semantic similarity values. The algorithm regarding processing of resource query in the proposed model takes an input query with the information of total Gridlets and total node resources. The output is the submitted Gridlet that best matched the node resource. There are two main loops of the algorithm - upper and inner. The upper loop runs for each Gridlet and the inner loop runs for each resource to find out the possible matched resources for Gridlet. Once a node is matched, it adds the matched node in the list and also adds its proximity and semantic values in corresponding lists. When all the three lists are ready for each Gridlet, then it calls the proposed model algorithm to pick the best matched node for the user Gridlet.

Routing Mechanism

The paper utilizes the Pastry overlay protocol (Rowstron & Druschel, 2001) for the routing mechanism in the proposed model. The protocol is scalable and self-administered and was initially designed for wide area P2P applications such as global data sharing & storage, group communication and naming purpose. However, we utilize this protocol in a Grid environment as a Grid system resembles a P2P architecture in the context of resource-sharing environments (Karaoglanoglou & Karatza, 2009). The main reason for choosing the Pastry over other P2P structured protocols is that the Pastry reduces communication overheads compared to others

such as Chord (Stoica et al., 2003). A detailed comparison of structured P2P protocol has been presented in (Lua, Crowcroft, Pias, Sharma, & Lim, 2005). First, we build decentralized overlay network in the FreePastry simulator (Druschel et al., 2012) using the Pastry protocol. After creating the P2P overlay network, we develop nodes and insert Grid resources under these nodes. Grid resources are created in GridSim (Buyya & Murshed, 2002) and are integrated with FreePastry to carry out simulations in a decentralized fashion. Both simulators support discrete event and time based simulations. The reason for using two simulators in the proposed model is because GridSim toolkit provides only basic network characteristics, whereas the creation of the P2P overlays facilities is not available with existing APIs (Application Programming Interface). This is the reason we integrate a FreePastry simulator for managing network in the Grid. We use GridSim which provides the platform that is close to a real Grid environment for creating Grid resources related

parameters. On the other hand, FreePastry provides a decentralized overlay network with efficient routing and location mechanism. Along with the basic resource characteristics, we add two additional parameters in Grid resources such as the processor architecture and the operating system because these are basic requirements for the user jobs. The Pastry assigns a 128-bit identifier Node Id for each node in hexadecimal format. Each node maintains a leaf set, routing table, and a neighborhood set that carries the latest information about other nodes (Rowstron & Druschel, 2001) and keeps track of their immediate neighbors. As the routing mechanism is concerned, a node can route a message/query to its numerically closest nodes. In the Pastry, the total routing hops/steps are less than $(\log_B N)$ steps under normal operation (N is a number of nodes and $B = 2^b$ where $b =$ number of bits used for the base of the chosen identifier with a typical value 4). The routing mechanism can be seen in Figure: 3.

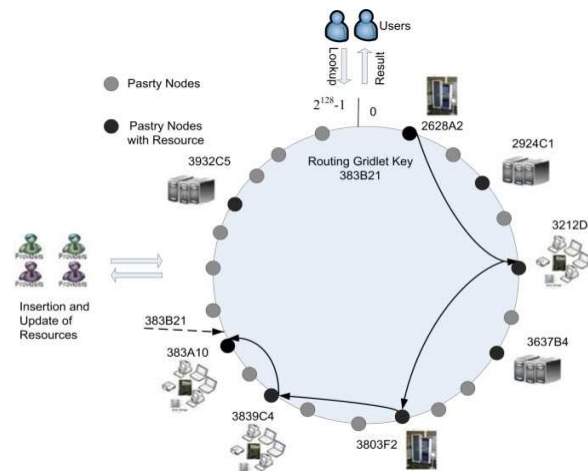


Figure: 3 Routing mechanism in Pastry logical ring

Figure: 3 shows that there are 22 Pastry nodes in the Pastry circular space where 8 nodes are occupied with Grid resources. Once a network and nodes are established, the providers publish resources through the insert method of the Pastry routing on the network and then randomly generated Ids are obtained against each resource node, which are shown in Figure:

3. Each Gridlet is sent to resource nodes to check whether available nodes can fulfill the requests or not. In the above scenario, Node Id 2628A2 routes the Gridlet Key 383B21 to the node closest to the value of the Gridlet key. It means the Gridlet with Key 383B21 is sent to the Node Id 383A10, which is numerically closest to the Gridlet's Key. When Node Id 2628A2

receives the Gridlet key *383B21*, then according to the Pastry routing algorithm, first, it checks the leafset entry of the node. However, in the above scenario, the node information does not fall in the leafset entries as both Gridlet's key and the node id are quite different. Based on the routing table information, at least that node sends the Gridlet to Node Id *3212D4* as prefix '3' is common in that node. Now Node Id *3212D4* will repeat the same routing process, and route to the Node Id *3803F2* where two common prefixes are '38'. In the same way, Node Id *3803F2* route this Gridlet to Node Id *3839C4* where three prefixes '383' are common. Finally, the Node Id *3839C4* finds the entry of the destination node in its upper range of leafset. In this way, a query can efficiently route in a highly distributed network with a minimum number of hops. As in the above scenario, the Gridlet finds a target Node Id *383A10* within four hops. When the Gridlet reaches the destination node then the comparison process starts. In case an exact match is not available for the Gridlet, then it will go for semantic matchmaking based on the semantic threshold value set by the users. If it matches the semantic threshold value with the resource's semantic similarity value, then the resource is considered as a matched resource otherwise the Gridlet will be rejected from this resource and it will move forward to another free resource. Details about measuring the semantic similarity are discussed in Sub-section 3.3 and the proximity model is described in Sub-section 3.4. By using the Pastry protocol with a combination of proximity and semantic data, we can get the optimal resource from the existing matched resources along with the efficient routing process. Moreover; the sub-domain based ontology structure enhances the job success rate in the network because it avoids the selection of irrelevant resources. In the next section, we describe the semantic mapping in a decentralized resource discovery model.

Semantic Mapping

The semantic mapping of ontologies in resource discovery services for a Grid computing is explained here. Grid resources belong to different virtual organizations with their own rules and policies, so it is possible for the same

resources to be published with different terminology. A semantic approach can be useful to identify the relationship between those resources (Chen & Tao, 2008). Ontologies can improve the quality of information and facilitate the increase in the efficiency of resource management in a Grid system (Vidal, Jos, Silva, Kofuji, & Kon, 2007).

Different from the traditional domain-based ontologies structure, we present the sub-domain ontologies structure to avoid the selection of non-relevant resources for the allocating of Grid resources. Towards this end, we extend and develop two ontologies of *Processor Architecture* and *Operating System* using the Protégé software (Standford, 2011). The ontologies help in finding relevant semantic resources in case the exact match is missing for job requirements and reduces the job rejection rate.

We compute semantic similarity values among concepts of ontologies. Semantic similarity is defined as the relationship between ontology concepts. The similarity of concepts represents the degree of commonality between these concepts. No standard procedure is available to measure the semantic similarity. However, a survey paper (Schwering, 2008) compares and contrasts the various models to measure the semantic similarity distance between ontology concepts. In paper (Schwering, 2008), the authors state that the selection of the measurement process is extremely complicated for certain applications as the human similarity judgment process is varied from person to person based on context and experience. For our implementation, we select the semantic measurement equation based on the network model because Network models measure similarity based on the notion of the distance short path algorithm. The Network model based semantic measurement technique has been proposed in (Andreasen, Bulskov, & Knappe, 2003) and also used in a decentralized semantic resource discovery model (Liangxiu & Berry, 2008). The authors derive conceptual similarity using the notion of "similarity graph". In this, the ontology is represented as a graph with concepts as nodes and relationships connecting these concepts as edges. The ontology of Grid

resources such as Processor ontology and Operating system can be referred to in our earlier paper (Shaikh, Alhashmi, & Parthiban, 2012). The advantage of the sub-domain based ontologies structure is that, only the relevant sub-ontology will be targeted by the query instead of the whole ontology. In this way, there is no chance to pick irrelevant resources as that could happen in a domain-based ontology structure.

Calculation of Semantic Similarities values

We present the method to compute semantic similarity values between concepts of Grid

resources ontologies. The equation (1) has been used in (Andreasen et al., 2003) to measure the degree of semantic similarity that uses a similarity function between concepts of ontologies. We utilize the equation (1) to calculate semantic similarity values between concepts of the above-mentioned ontologies. The measurement of semantic similarity values has been done among concepts of developed ontologies that represent the degree of commonality between concepts. This commonality shows how concepts are semantically relevant to each other. It is known as semantic similarity values and denoted by Ψ symbol

$$\Psi(x, y) = \rho \frac{|\alpha(x) \cap \alpha(y)|}{|\alpha(x)|} + (1 - \rho) \frac{|\alpha(x) \cap \alpha(y)|}{|\alpha(y)|} \quad (1)$$

In (1), ρ is a factor that determines the degree of influence of generalization of ontology concepts. The value of ρ lies between 0 and 1. If the value of ρ is 1, that means perfect generalization, with each and every concept defined properly and 0 means extremely poor generalization. We set the different ρ value such as 0.25, 0.5, and 0.75. However, the results show in this paper by using ρ value 0.50.

In the equation 1, $\alpha(x)$ is the set of nodes reachable from x and $\alpha(x) \cap \alpha(y)$ the reachable nodes shared by x and y . $(x, y) = 0$ means x and y are entirely dissimilar and $\Psi(x, y) = 1$ means full similarity. Table 2 shows the semantic similarity values for partial *Processor Architecture* ontology with $\rho = 0.50$.

The purpose of offering semantic resources to users is to reduce the job rejection in a Grid system when an exact match is not available. However, if a Gridlet has any specific requirements or has resource compatibility issues, then semantic resources could not fulfill the user requests properly and as a result, Gridlets can fail at run time. To get the maximum benefits from a semantic decentralized resource discovery model, it is assumed that all Gridlets are considered cross platform where resource compatibility is not an issue.

Selection and Matching of optimal resources in proposed Model

This section explains the process of the selection of optimal resources among Grid resource in the proposed model. We merge the proximity of nodes and semantic similarity values and utilize in a semantic sub-domain based decentralized resource discovery model. The purpose of this unification is to get optimized resources for user jobs, so that the Grid brokers could select optimal resources in terms of proximity with high semantic relevant resources. This optimization model improves traditional selection mechanism and picks the optimal resources. It could also be used in economic Grid (Shaikh, Alhashmi, & Parthiban, 2013) as the utilization of optimal resources can make better profit compared to conventional resource provisioning mechanisms. In this optimization model, we present the unification of proximity and semantic similarity values based on a ranking method. By doing so, we get all possible matched nodes for a current Gridlet based on the semantic similarity values then compute the proximity between nodes. Proximity distance is measured through a scalar proximity metric in the Pastry overlay, which is based on IP routing hops. After normalizing the values, we select the optimized resources for

user jobs. By applying the model, the application performance can be improved.

Experiment configuration and results

This section discusses the experiment and its results.

Experiment

Two discrete event based simulators i.e. GridSim and FreePastry are integrated to measure the

efficiency of both Grid entities and network related performance metrics. The proposed model deployed the algorithm in the sub-domain ontology structure to improve the recall values. The proposed model uses the unification of proximity and semantic similarities values in the selection process of resources for user jobs. The performance of the model is highly dependent on the number of ontology concepts and the semantic threshold values set by users. We have run the simulations with the following set of parameters. The experiment configuration is shown in the following table:

Table 3: Experiment Configurations

Simulation Parameters	Relevancy	Values
Number of Nodes	Providers	1024
Number of Resources		1024
Number of Machines per resource	Providers	4-10
Number of Processors per resource	Providers	5-80
Rating of PE (MIPS)	Providers	350-450
Semantic Threshold (τ)	Users	0.25
Degree of Generalization (β)	Providers	0.50
Network Bandwidth (MB)	Providers	100
Number of Gridlets/Jobs/Queries	Users	250
Processor Architecture ontology	Providers/Users	53
Operating System (OS) ontology	Providers/Users	39
Time To Live (TTL)	Users	20
Performance Metrics	Providers/Users	Proximity, Semantic similarity

Table 3 shows that the parameters and their values along with their relevancy either with users or providers are used in our experiment. The parameter values of the Grid entities that are shown in Table 3 are generated using Random Uniform Distribution. The reason of using Random Uniform Distribution is that it is effectively distributed according to the standard uniform distribution and useful to run simulation experiment. The aim of this simulation is to measure the proximity and

semantic similarity and compare the results with FCFS scheduling. The details of experimental results are as follows:

Experimental Results

The results and comparison of the optimized proposed model with the existing semantic decentralized resource discovery model are explained as follows:

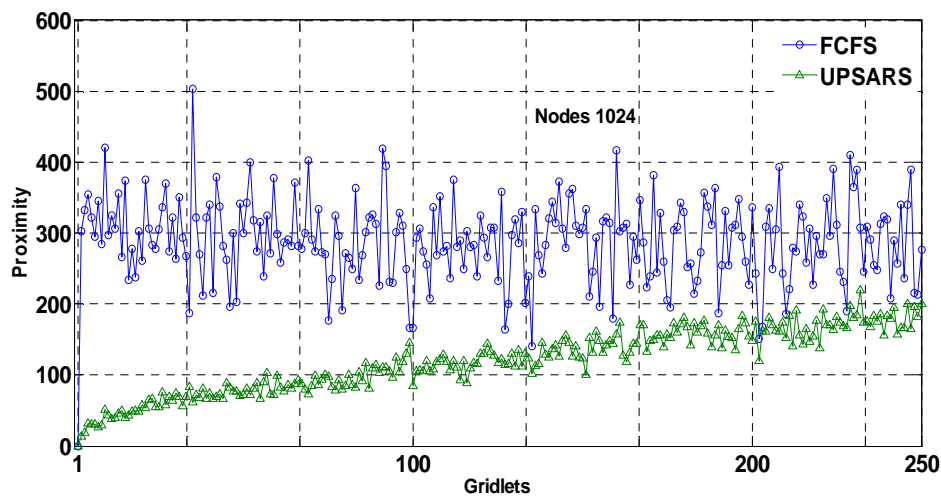


Figure 4 : Comparison of Proximity between the proposed model and FCFS

Figure 4 illustrates the comparison of proximity between the proposed model and FCFS. We query about 250 Gridlets under 1024 nodes. The graph shows that most of the Gridlets in the FCFS are scheduled farther than the proposed model. For example, in the above graph, the 100th Gridlet is scheduled at the resource that has a proximity value around 90 in the proposed model; whereas the same Gridlet is scheduled in FCFS that has a proximity value of 150. Because

the proximity factor is not being considered in FCFS scheme as the selection of Grid resources is based on the semantic matched node, it is highly probable to schedule the jobs anywhere in the network. However, the proposed model outperforms FCFS in terms of proximity as it utilizes the proposed model algorithm where Gridlets are scheduled on nearby nodes. In the proposed model, the resources are allocated to the user closest nodes that can enhance the Grid performance.

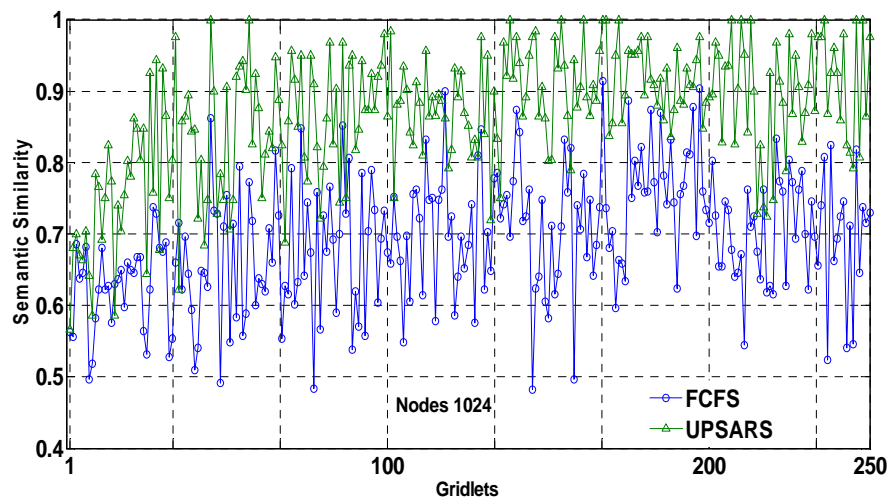


Figure 5: Comparison of semantic similarity between proposed model and FCFS

Figure 5 shows that most of the Gridlets are scheduled on high semantic relevant resources in the proposed model as compared to FCFS scheduling. In FCFS, users have a low chance to get a high degree of relevant resources as compared to the proposed model even when the best semantic similarity resources are available. However, in the proposed model, most of the time, users get high semantic relevant resources. Each Gridlet has its own requirement and we inject the same type of requirement of Gridlet in both models to fair comparison. It is noted that there is no proportion between semantic similarity and Gridlet value so the increase in number of Gridlets cannot affect semantic similarity values.

Conclusion and Future work

In this paper, a novel optimization model has been presented that selects optimal Grid resources for scheduling user jobs by considering proximity and semantic similarities values. The model is designed and implemented when a gap is identified in an existing FCFS allocation scheme for a semantic decentralized resource discovery. To overcome the gap, the proposed model utilizes the best combination of proximity and semantic values of available Grid resources and enhances the Grid performance. The experimental results verified that the proposed model provides benefits in the allocations of most suitable resources. The experimental results are compared with the existing FCFS scheduling that shows that the proposed model outperforms in terms of proximity and semantic similarity. In the future work, we would like to extend the ontology of Grid resources and implement and deploy the proposed model in real Grid system with real world applications.

References

1. Andreasen, T., Bulskov, H., & Knappe, R. (2003). *From ontology over similarity to query evaluation*. Paper presented at the 2nd CologNET-ElsNET Symposium-Questions and Answer: Theoretical and Applied Perspective, Amsterdam, Holland.
2. Buyya, R., & Murshed, M. (2002). GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for Grid computing. *Concurrency and Computation: Practice and Experience*, 14(13-15), 1175-1220.
3. Chen, L., & Tao, F. (2008). An Intelligent Recommender System for Web Resource Discovery and Selection *Intelligent Decision and Policy Making Support Systems* (pp. 113-140).
4. Druschel, P., Haeberlen, A., Hoyer, J., Iyer, S., Mislove, A., Nandi, A., . . . Singh, A. (2012). Free Pastry Software Retrieved Jan 20, 2012, from <http://www.freepastry.org/FreePastry/>
5. Iamnitchi, A., Foster, I., & Nurmi, D. C. (2003). A peer-to-peer approach to resource location in grid environments *INTERNATIONAL SERIES IN OPERATIONS RESEARCH AND MANAGEMENT SCIENCE* (pp. 413-430): Springer.
6. Karaoglanoglou, K. I., & Karatza, H. D. (2009). *Performance evaluation of a resource discovery scheme in a Grid environment prone to resource failures*. Paper presented at the Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on.
7. Li, J. (2010). Grid resource discovery based on semantically linked virtual organizations. *Future Generation Computer Systems*, 26(3), 361-373.
8. Liangxiu, H., & Berry, D. (2008). Semantic-Supported and Agent-Based Decentralized Grid Resource Discovery. *Future Generation Computer Systems*, 24(8), 806-812.
9. Lua, E. K., Crowcroft, J., Pias, M., Sharma, R., & Lim, S. (2005). A survey and comparison of peer-to-peer overlay network schemes. *IEEE Communications Surveys and Tutorials*, 7(2), 72-93.
10. Pirrò, G., Talia, D., & Trunfio, P. (2012). A DHT-based semantic overlay network for service discovery. *Future Generation Computer Systems*, 28(4), 689-707.

11. Qureshi, M. B., Dehnavi, M. M., Min-Allah, N., Qureshi, M. S., Hussain, H., Rentifis, I., . . . Xu, C.-Z. (2014). Survey on Grid Resource Allocation Mechanisms. *Journal of Grid Computing*, 1-43.
12. Ranjan, R., & Buyya, R. (2009). Decentralized overlay for federation of Enterprise Clouds. *Handbook of Research on Scalable Computing Technologies*, 191.
13. Ranjan, R., & Buyya, R. (2009). *Decentralized Overlay for Federation of Enterprise Clouds, Handbook of Research on Scalable Computing Technologies*. USA IGI Global.
14. Rowstron, A., & Druschel, P. (2001). *Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems*.
15. Schwering, A. (2008). Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. *Transactions in GIS*, 12(1), 5-29.
16. Shaikh, A., Alhashmi, S., & Parthiban, R. (2012). A Semantic Impact in Decentralized Resource Discovery Mechanism for Grid Computing Environments. In Y. Xiang, I. Stojmenovic, B. O. Apduhan, G. Wang, K. Nakano & A. Zomaya (Eds.), *Algorithms and Architectures for Parallel Processing* (Vol. 7440, pp. 206-216): Springer Berlin Heidelberg.
17. Shaikh, A., Alhashmi, S., & Parthiban, R. (2013). *Ontology-based Decentralized Resource Provisioning in Economic Grids*. Paper presented at the Proc. 20th International Conference on Business Information Management Association Kuala Lumpur, Malaysia.
18. Somasundaram, T., Govindarajan, K., Kiruthika, U., & Buyya, R. (2014). Semantic-enabled CARE Resource Broker (SeCRB) for managing grid and cloud environment. *The Journal of Supercomputing*, 68(2), 509-556. doi: 10.1007/s11227-013-1047-z.
19. Stanford. (2011). Ontology Editor & Knowledge base Framework. Retrieved 12 Dec 2011, from <http://protege.stanford.edu/>.
20. Stoica, I., Morris, R., Liben-Nowell, D., Karger, D. R., Kaashoek, M. F., Dabek, F., & Balakrishnan, H. (2003). Chord: a scalable peer-to-peer lookup protocol for internet applications. *Networking, IEEE/ACM Transactions on*, 11(1), 17-32.
21. Tam, D., Azimi, R., & Jacobsen, H. A. (2004). Building content-based publish/subscribe systems with distributed hash tables. *Databases, Information Systems, and Peer-to-Peer Computing*, 138-152.
22. Tao, Y., Jin, H., Wu, S., & Shi, X. (2009). Scalable DHT- and ontology-based information service for large-scale grids. *Future Generation Computer Systems*, 26(5), 729-739.
23. Vidal, A. C. T., Jos, F., Silva, d. S. e., Kofuji, S. T., & Kon, F. (2007). *Semantics-based grid resource management*. Paper presented at the 5th international workshop on Middleware for grid computing, Newport Beach, California.