Lightweight Blackboard Resource Discovery Mechanism
For Unstructured Peer To Peer Networks

Azrul Amri Jamal and William John Teahan

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Malaysia
School of Computer Science, Bangor University, United Kingdom

Abstract: Grid computing relies heavily on resource discovery techniques in order to find shared resources in the network. Resource discovery techniques can be categorised into two: informed; and uninformed. Blackboard resource discovery mechanism (BRDM), an informed resource discovery technique utilised in grid computing has been examined in this research. Three shortcomings of BRDM and its previous experiments have been outlined and discussed. The technique is then improved by reducing the network cost of the BRDM. The improved BRDM, lightweight BRDM (LBRDM), is then analysed on simulated unstructured peer-to-peer networks in PeerSim simulator. The technique is assessed using combined query efficiency metric that includes network communication cost and successful searches. LBRDM has shown significant improvement in combined query efficiency compared to BRDM in all parameters evaluated.

Key words: Resource Discovery • Lightweight • Informed Search • Unstructured • P2P • BRDM • Grid Computing

INTRODUCTION

Foster et al. first coined grid computing in 1990s and outlined coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organisations as the main challenge in grid computing [1]. Resource discovery plays a significant role in order for grid computing to achieve coordinated resource sharing [2]. Resource discovery techniques that being used should be able to search for resources in peer-to-peer (P2P) networks in efficient manner. Early implementation of resource discovery techniques usually flood the network with search queries. This usually results to undesirable effects such as wasting computing and networking resources and degeneration of communication networks [2].

Grid computing can be implemented in various kind of network such as: centralised, decentralised, P2P, hierarchical and agent based system. P2P can be categorised into four types, namely unstructured, structured, super-peer and hybrid [3]. Resource discovery in unstructured P2P network rely more on flooding the network compared to three other P2P network categories. This is due to the fact that it does not have a clear structure and it also lacks super-peers where there are some nodes are treated as better than others. Resource discovery in structured network usually rely on the node at the top of the structure, while in super-peer network, resource discovery can direct their search towards super-peers where there are a lot of other nodes connected to them [2].

Resource Discovery: As discussed by [2], resource discovery techniques can be divided into two categories: informed and uninformed. Informed search techniques utilise some heuristic approach towards finding the resources, whereas uninformed search techniques do not [2,4]. The reliance on heuristic has made the informed resource discovery techniques to have better query efficiency when compared to uninformed resource discovery techniques [2].

Resource discovery techniques classified into the uninformed category are: Random Walk; Restricted Random Walk; Breadth-First Search (BFS); Depth-First Search; and Alpha Breadth-First Search (α-BFS), while techniques that fell into the informed category are: Intelligent-BFS (Int-BFS); Adaptive Probabilistic Search (APS); Blackboard Resource Discovery Mechanism.
(BRDM); Learning Automata-based Resource Discovery (LARD); Improved APS; and Discovery of Heterogeneous Multiple Compute Resources Framework (DHMCF), Distributed Resource Location Protocol (DRLP) [2,5].

Blackboard Resource Discovery Mechanism: BRDM was first coined by [6], it is an informed resource discovery technique that adopted the blackboard mechanism from the field of artificial intelligence [6]. The technique is utilised in ParCop grid computing to find available resources in the network [7]. Each node that has been implemented with BRDM will have two blackboards, where one contains list of recommendable adjacent neighbours \((N_1)\) while the other blackboard list non-recommendable adjacent neighbours \((N_2)\).

The recommendations are based on whether or not the neighbouring nodes forwarded the query message that has been forwarded to them [5]. The information recorded in both blackboards will be used by BRDM to decide where to forward all of future queries. By limiting query forwarding (QF), BRDM was able to reduce network cost and increase its query efficiency. [6] has shown the simulation method that has been done to evaluate BRDM. The technique was proven to return better results than several other techniques, such as: Int-BFS; DRLP; BFS; APS; and RW. Nevertheless, there are three issues that need to be addressed in BRDM simulations: small simulation environment; non-recommended list type I error; and high network cost for unsuccessful searches.

Small Simulation Environment: Simulation done in [6] was done on nodes that are placed in 100 by 100 plots. Based on the number of plots shows that the maximum number of nodes being tested is 10,000. This limit can be considered as small, considering that real implementation of unstructured P2P networks would have involved millions of nodes. The small number of nodes used in previous BRDM simulations might due to the limitation in computing power and memory at the time the work was published.

Non-Recommended List Type I Error: Forwarding queries to adjacent neighbours listed in \(N_1\) and avoiding QF to adjacent neighbours listed in \(N_2\) were able to reduce the number of QF and increase BRDM’s query efficiency. Nonetheless, in the early stages of message forwarding by BRDM, not all neighbours are forwarded with queries. Thus, the neighbouring node will be included in the \(N_2\) list and it won’t receive any future queries even though resources might be available near to the node. Being listed in the non-recommended list is called as Type I error.

High Network Cost for Unsuccessful Searches: There are several metrics can be used to quantify resource discovery techniques. Among others are number of query forwarded, number of successful searches and query efficiency. Query efficiency was introduced by [2], the metric take information such as number of query forwarded, number of successful searches and number of nodes in the network into calculations [2]. Query efficiency is denoted as \(\eta\) (common Greek letter to show efficiency), the equation of \(\eta\) is as follows:

\[
\eta = \frac{\text{QueryHits}}{\text{QueryMessagesForward}}
\]

where \(N\) is the number of nodes in the whole network being searched [2].

Nonetheless, the metrics shown above only take the network cost of query message forwarding, \(QF = \prod^TTL \{l(n_o)\}\), where \(l(n_o)\) is the number of query forwarded from the originating node and \(TTL\) is the time-to-live of the query. The cost of network resources being used to return queries are not taken into account. In BRDM, feedbacks for all successful and unsuccessful searches are returned to the originator using the same path that were used to forward the queries. Therefore, the cost for the network resources used is actually double than calculated above, at \(2\prod^TTL \{\{l(n_o)\}\}\). A new efficiency evaluation technique is needed to calculate the efficiency of BRDM. The new metrics, \(\eta^*\), calculates the effectiveness with consideration of query message forwarding and their feedbacks. The equation for \(\eta^*\) are as follows:

\[
\eta^* = \frac{\text{QueryHits}}{\text{QueryMessagesForward} + \text{QueryMessagesFeedback}}
\]

\[
= \frac{\text{QueryHits}}{\prod^TTL \{l(n_o)\} + B_s + B_u}
\]

where \(B_s\) and \(B_u\) are the number of successful and unsuccessful searches respectively.
Table 1: Five-tuple Alpha Multipliers

<table>
<thead>
<tr>
<th>Alpha Multipliers Sets</th>
<th>Five-tuple Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set B</td>
<td>{1.0, 0.8, 0.6, 0.4, 0.2}</td>
</tr>
<tr>
<td>Set M</td>
<td>{0.5, 0.5, 0.5, 0.5, 0.5}</td>
</tr>
<tr>
<td>Set N</td>
<td>{0.4, 0.4, 0.4, 0.4, 0.4}</td>
</tr>
</tbody>
</table>

Based on the shortcomings of the BRDM and its previous experiments, the resource discovery technique will be simulated in a larger simulation environment. The selected environment should be high in scalability and is able to handle at the very least one million simulated nodes. BRDM’s non-recommended list \((N/2)\) on each node are not used for resource discovery in order to eliminate the non-recommended list type I error and high network cost of successful searches.

**Methodology:** [2] discussed three methods to examine a search algorithm, they are: mathematical; using simulations; and experiments on actual P2P systems. Proofing using mathematical models are based on assumptions and simplification of the algorithm and the network. Experiments using actual P2P systems are difficult to run and they are usually costly and require significant amount of resources. All of the shortcomings of the mathematical model and actual P2P test bed can be avoided by using simulations [2,5].

Several P2P simulators have been examined and compared in [2], namely: 3LS; General Peer-to-Peer Simulator; Neurogrid; P2PSim; PeerSim; PeerThing; Query Cycle; and Real Peer. It is found that PeerSim has the best overall marks based on its usability; documentation; scalability; extensibility; runtime; status; and GUI. The unstructured P2P network selected are generated using the scale free Barabási-Albert topology \((k = 2)\). All of the experiments are done in 20 cycles and the search query has a TTL of 20.

The one million simulated nodes are connected with undirected connection. The nodes and neighbours’ distribution uses the random seed of 1234567890, 1415926535 and 8979323846. Table 1 shows the five-tuple alpha multipliers that being used for this research. The sets, Set B, Set M and Set N, are obtained from [2], where the alpha multipliers came out with the best search results for \(\alpha\)-BFS with 63.15%, 62.78% and 62.23% of combined efficiency respectively.

The improved BRDM is lightweight in terms of the network resource that it uses compared to BRDM. Thus, the technique is called Lightweight Blackboard Resource Discovery Mechanism, LBRDM. LBRDM will be compared to BRDM using the metrics, \(\eta^*\) and successful searches \((ss)\), that has been discussed in previous section. The combination of \(\eta^*\) and \(ss\) is called as Combined Query Efficiency, \([(\eta^* + ss)/2]\)%, where the metric take into consideration the query efficiency and the number of successful searches that the technique achieved.

![Fig. 1: Combined Query Efficiency Star \([(\eta^* + ss)/2]\) \((N=1\ million, k=2,\ TTL=20)\)
Table 2: Average Query Efficiencies (\(\eta\) and \(\eta^*\)) and Successful Searches (ss)

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Alpha Multipliers</th>
<th>(\eta&amp;\eta^*)</th>
<th>(\eta^*&amp;ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRDM</td>
<td>Set B</td>
<td>89.30</td>
<td>70.65</td>
</tr>
<tr>
<td></td>
<td>Set M</td>
<td>80.40</td>
<td>62.84</td>
</tr>
<tr>
<td></td>
<td>Set N</td>
<td>68.15</td>
<td>47.60</td>
</tr>
<tr>
<td>LBRDM</td>
<td>Set B</td>
<td>84.12</td>
<td>86.31</td>
</tr>
<tr>
<td></td>
<td>Set M</td>
<td>84.01</td>
<td>86.28</td>
</tr>
<tr>
<td></td>
<td>Set N</td>
<td>69.02</td>
<td>70.86</td>
</tr>
</tbody>
</table>

**RESULTS AND DISCUSSION**

BRDM and LBRDM are tested on three different random seeds and three different sets of alpha multipliers. The results are then inserted into \(\eta\) and \(\eta^*\) and combined with the percentage of successful searches, ss. Table 2 shows the averaged results of each technique on all alpha multiplier sets. LBRDM shows better results compared to all of its BRDM counterparts. Figure 1 shows the graph of the Combined Query Efficiency Star, \((\eta^* + ss)/2\). By eliminating the requirements for non-recommended list, LBRDM technique was able to significantly reduce the network communication cost while maintaining high successful searches.

This research has proven to reduce network communication cost for BRDM, a resource discovery mechanism that is used in grid computing. The reduction of communication, traversal and storage of search information did not hinder LBRDM to produce a better combined query efficiency compared to BRDM. Further LBRDM research can be done by increasing the number of simulated nodes, implementation on real P2P systems and experiments on other network characteristics.

**ACKNOWLEDGEMENTS**

Authors would like to thank Universiti Sultan Zainal Abidin (UniSZA) and Bangor University for funding and preparing an environment conducive to this research.

**REFERENCES**