Incentive Driven Node Discovery in a P2P Network Using Mobile Intelligent Agents

Prithviraj(Raj) Dasgupta
Department of Computer Science
University of Nebraska, Omaha, NE 68182.
E-mail: pdasgupta@mail.unomaha.edu

Abstract

A peer-to-peer(P2P) network consists of nodes that are capable of sharing resources and services with each other in a decentralized and distributed manner. A node joining a P2P network uses the node discovery protocol to discover other nodes already in the network. A node then uses the resource discovery protocol to distribute and acquire resources from other nodes in the network. Existing algorithms for P2P networks optimize the resource discovery protocol to guarantee rapid and reliable resource sharing. In this paper, we describe an intelligent node discovery protocol using a mobile agent enabled P2P system. Our simulation results show that agent based node discovery techniques in P2P systems can compare to and even outperform traditional node discovery techniques.

Keywords: Mobile agents, P2P networks, node discovery, gossip based learning.

1 Introduction

The peer-to-peer(P2P) model[9] provides a suitable paradigm for agents to interact and cooperate with each other. P2P systems consist of a network of peer nodes that are capable of sharing resources and services with each other. A P2P network is dynamically created and destroyed as peer nodes join and leave the network. Therefore, the topology of a P2P network is determined largely by the incentives to the users at the peer nodes to share and distribute resources. These incentives comprise several factors including the connectivity of the nodes in the network that the user wishes to join, availability of resources at those nodes and the user's requirements. Current methods for discovering nodes in a P2P network include naive techniques such as flooding the network[6] and refined algorithms such as local indexed based node discovery [16] and rumor mongering (gossip)[11].

We envisage that a significant performance improvement can be achieved in discovering nodes in P2P networks by using mobile intelligent agents that dynamically determine the network topology. Experimental simulations of the node discovery protocol using our mobile agent based P2P system indicate that agent based techniques compare favorably with traditional node discovery techniques in P2P networks.

2 Peer-to-Peer (P2P) Network Protocols

A P2P system is a decentralized and distributed network of nodes that is capable of sharing and distributing resources between themselves. The primary objective of a node in a P2P network is to search and acquire resources and services available on other nodes in the network and, simultaneously allow other nodes to access resources present on the node itself. For this every node uses a node discovery protocol to determine other nodes in the network and a resource discovery protocol to determine resources on other peer nodes. The node discovery protocol consists of two messages. A node uses a ping message to probe and discover other nodes in the network. A pong message is a response to a ping message and it contains information about the replying node including the Internet address, available bandwidth, and networking resources. After a node becomes aware of its peer nodes it uses the resource discovery protocol to discover resources that are possibly present on those peers. The resource discovery protocol comprises the query message that is forwarded to successive peers until the resource is discovered or the lifetime of the message expires. If the resource is found on a peer node a queryHit message is sent back to the node originating the query. The requesting node and the providing node then decide on the exchange protocol for the resource.

2.1 Incentive Driven Node Discovery

The nodes in a P2P network join and leave the network in an ad-hoc manner. A node joins the P2P network when the user at the node either wishes to download resources from users on other peer nodes in the network, or, desires to share resources with other users. In both situations, the decision of the user to join the P2P network is driven by an incentive based on the availability of resources. The objective of a user joining a P2P network is always to connect to peer nodes that promise the best incentives from the resulting association.

The incentive driven model can be also applied to node discovery in the P2P network. The incentive for the user node would be to connect to nodes that are potentially resourceful. At the network level, resourceful nodes are characterized by the following parameters:

- **Node Availability:** A highly available node is one that remains accessible for a considerable time and also responds reasonably quickly to information queries. Highly available nodes are suitable candidates to connect to in a P2P network because they might have acquired substantial resources over time. Moreover, a reasonable response time guarantess that unsuccessful resource queries do not delay other users.

- **Node Connectivity:** A well-connected node is likely to interact with several nodes in the network and can provide information about other suitable nodes to connect to.

- **Reliability of Information:** The information provided by a node cannot be guaranteed to be authentic. Reliability of the information exchanged in a P2P network is not addressed in the traditional P2P model[9]. In our model, we have accomodated information reliablity by appending the information with a reliability value. It can be argued that this reliability value can be manipulated by malicious nodes. However, during node discovery, information about other nodes in the network and the reliability of that information is continuously updated as the network discovery protocol proceeds. Dynamic update of the reliability information of a node from different nodes can reduce collusions to malign the reputaion of a node.

A node joining the P2P network should selectively connect to nodes that possess the best values for the above parameters. During node discovery, the parameters already obtained from a node can change even as other nodes are getting discovered. Therefore, the information about other nodes discovered by a node should be continuously updated. Mobile agents can be suitably used to determine the network related parameters of the nodes in a P2P network, select a node based on the incentives offered to the user and dynamically update the network and node information that has already been collected. In the next section we describe an agent based framework for dynamically discovering nodes that offer best incentives for a new node to connect to, using a gossip based learning mechanism.

3 Mobile Agent Based Node Discovery in P2P Networks

A mobile agent is a piece of software code that can be sent across a network to a remote location, and continue its execution after it reaches the remote site. A mobile agent that is intelligent can also react to changes in the environment, learn from the environment and use the information learnt to refine its operation so that it can perform its tasks more efficiently. Mobile intelligent agents can be used to implement the node discovery protocol in P2P networks by adaptively learning about suitable nodes to visit while migrating through different nodes in the network.

3.1 Agent Framework

To support the node discovery protocol in a P2P system, every node contains the agents shown in Figure 1.

We have used the Java based IBM Aglets SDK [8] to implement the agent framework. In the Aglets SDK, agents called *aglets* are implemented as Java threads that can be serialized into a TCP/IP message and sent to a remote location. On arrival at the remote location, the thread is deserialized and the aglet can perform the task designated for that location. Aglets are executed inside an aglet server. Therefore, each node in our P2P network should run an aglet server so that it can send and receive aglets. In our system, aglets communicate with each other using asynchronous messages in a language similar to FIPA ACL[3]. Aglets are also capable of remote messaging. However, we have restricted aglets to communicate locally with other aglets within the same server to ensure communication security and improve congestion.

4 Node Discovery Protocol

The node discovery protocol is implemented in our system by the reconnaissance agent. The aglet server running on a node contains a controller agent that receives the user requests. When a user wishes to join the P2P network, he sends a join request to the controller agent. To join the P2P network, the controller agent requires the address of at least one more node from which the address requests can be further forwarded. We have used a central server model similar to Napster[10] to enable a node discover at least one neighbor. A central server contains the address of well connected nodes in the network. The addresses of nodes within the central server get updated as nodes join and leave the network.

The controller agent requests the central server to respond with the address of a well-connected node. The controller agent then creates a reconnaissance agent with the following attributes:

1. *unvisitedURLList* that initially contains the address of the well-connected node returned by the central server. The *unvisitedURLList* gets updated as the reconnaissance agent visits different nodes,

2. *numConn*, the desired number of nodes that the user wishes to have as immediate neighbors, and,

3. an incentive function used by the reconnaissance agent to determine the suitability of a node that it visits to be an immediate neighbor of its origin node

4.1 Agent Interaction Protocol

The reconnaissance agent visits each site in its *unvisitedURLList* and interacts with the information agent using the following protocol:

1. On arriving at a remote site, the reconnaissance agent submits the address of its origin node to the information agent.

2. The information agent responds with a communication key that is used to uniquely identify all communication the information agent has with the current reconnaissance agent.

3. The reconnaissance agent and information agent exchange information about addresses of other nodes that they are aware

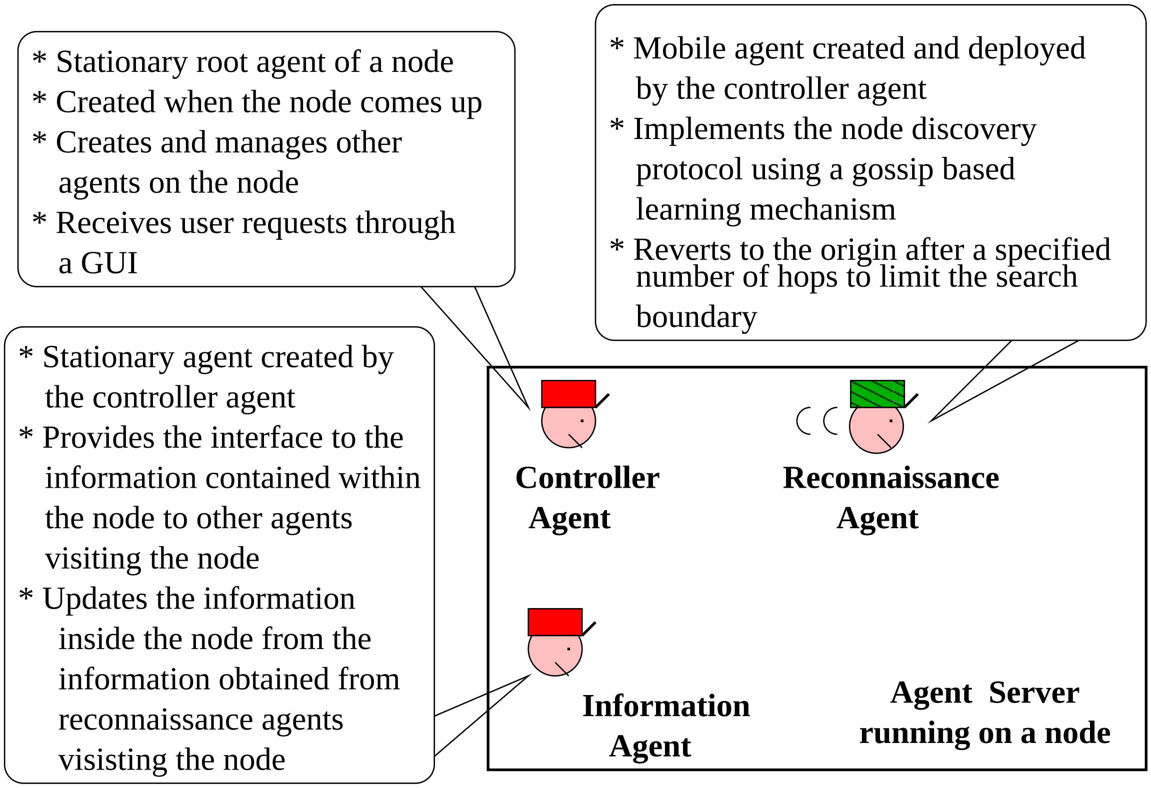

Figure 1: Different types of agents inside a node of an agent enabled P2P system for implementing the node discovery protocol.

of using a gossiping algorithm. Each address is accompanied by a score based on an incentive function and the reliability of the score.

4. The reconnaissance agent uses an incentive function to calculate a score for the node being currently visited. Simultaneously, the information agent on the node also uses its incentive function to determine a score for the origin node of the reconnaissance agent based on its interactions with the reconnaissance agent.

5. The reconnaissance agent updates the *unvisitedURLList* to include the addresses learnt from the information agent and ranks the addresses based on their score and score reliability.

6. The reconnaissance agent then proceeds to visit the nodes contained in the *unvisitedURLList* and repeats the same protcol on arriving at each node.

Address	Score	Reliability of Score

Figure 2: Gossip format exchanged during the agent interaction protocol.

7. When the reconnaissance agent has visited *numConn* nodes it reverts back to the origin and submits the list of addresses to the controller agent.

The controller agent selects the nodes returned by the reconnaissance agent that provide a high incentive to connect to based on the nodes' score and score-reliability.

4.2 Incentive Function

The gossip information format exchanged between the reconnaissance agent and information agent is shown in Figure 2. The address of a node in the P2P network is associated with a score for the node and a reliability of the score information. The score for a node is computed using the incentive function shown in Equation 2 when the reconnaissance agent from a node interacts with the information agent on a node it is visiting during step 4 of the agent interaction protocol described above.

$$
\begin{aligned}
&Score(node) = \\
&\quad w_1(\text{No. of successful communications-} \\
&\quad \text{with agent representing the node}) + \\
&\quad w_2(\text{No. of addresses obtained from agent}) \\
&\quad + w_3(\text{No. of previously undiscovered-} \\
&\quad \text{addresses obtained from agent})
\end{aligned} \tag{1}
$$

The terms on the rhs of Equation 1 are compatible with the incentive parameters of a node discussed in Section 2.1. The number of successful communications indicates that the agent and the node containing the agent have a reasonable response time and are available for communication with other nodes. The total number of addresses obtained from a node indicates the connectivity of the node, and, the number of new addresses discovered indicates that the node has been recently active and is aware of nodes that have recently joined the P2P network. The weights w_1, w_2 and w_3 indicate the preferences of the agent that is assessing the score of the node. For example, a high value of w_1 indicates that node availability is an incentive for the agent to attribute a high score, while high values for w_2 and w_3 indicate that the agent prefers highly connected and active nodes respectively.

The reliability of the score obtained in equation 2 is based on the inverse proximity of the source of the information. For example, when a reconnaissance agent and information agent interact and assign a score to each other, the score information is considered to be reliable. Each agent has directly interacted with the other and is aware of each others parameters on the rhs of Equation 1. On the other hand, the score information about another node that the reconnaissance agent receives from the information agent is considered less reliable because the reconnaissance agent has not had direct interaction with that node. Reliable information is assigned a low numeric value.

The reconnaissance agent and controller agent use a weighted sum of the score and score reliability obtained through gossip to rank nodes and select suitable nodes to visit or connect to as shown in Equation 2. w_r is a weight that the agents use to express their preference for reliable versus unreliable nodes.

$$
\begin{aligned}
&Rank(address) = \\
&\quad \text{Score of address} + \\
&\quad w_r * (\text{Reliability of score})
\end{aligned} \tag{2}
$$

5 Simulation Results

Our objective for the simulation runs was to verify the operation of our algorithm for incentive based node discovery. Therefore, we assumed that a P2P network with 10 nodes is already available and 3 of those nodes are well-connected. We assume that a well-connected node is a node with 3 or more active connections represented by gossip information that contains three addresses that have a highly reliable score. In all our simulation runs, we created a new node that wishes to join the P2P system. The desired number of connections *numConn* in the P2P network required by the joining node was set to 8. The agent preferences while determining the score of a node using Equation 1 was set to node availability ($w_1 = 1.0$) and node connectivity ($w_2 = 1.0$) with a moderate preference for discovering new nodes ($w_3 = 0.7$). The reconnaissance agent assigns a high weight $r = 0.8$ to the score reliability information while ranking the nodes using Equation 2. Results shown on the graphs were averaged over 5 simulation runs.

Figure 3(a) shows the number of actual active connections that the different nodes in the P2P network have, normalized over the maximum number of connections (9) that a node can have, and, the normalized score for the different nodes determined by the reconnaissance agent sent by the node that joins the

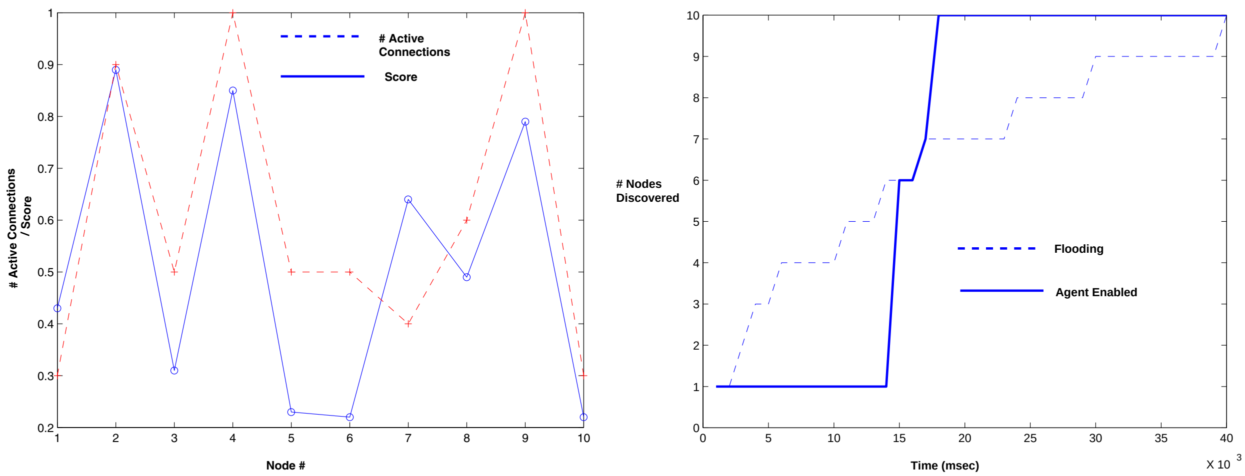

Figure 3: (a) Number of active connections of the nodes in a P2P network and the score of the node calcuated by the reconnaissance agent. (b) Time vs. the number of nodes discovered by flooding and the gossiping algorithm using mobile agents.

P2P network. The high degree of correlation between the actual number of connections and the score determined by the reconnnaissance agent shows that our algorithm is successful in determining nodes with high availability and connectivity and is suitable for node discovery.

Figure 3(b) compares the performance of our intelligent node discovery algorithm to the flooding technique used in P2P systems. Our experimental results show that the agent based technique is capable of detecting nodes more rapidly than flooding. The first node detected by the agent based node discovery technique is the well conneccted node returned by the central server. Thereafter, flooding initially detects more nodes because the reconnaissance agent expends time in migrating to remote nodes, gossiping and determining the score of a node. However, the reconnaissance agent returns to the origin node around 15 seconds through the simulation after discovering 6 out of the 10 nodes. Some of the nodes in the network were pendant nodes with only one connection and were therefore, difficult to reach by the reconnaissance agent. Although the reconnaissance agent obtained the information about these nodes through gossip, it did not visit the nodes because they had a low value for the reliability of the score information. Since the number of nodes discovered(6) was less than $numConn(8)$, the controller agent sent another reconnaissance agent to possibly discover more nodes and that agent returned in about another 5 seconds. By changing the value of r in Equation 2 from 0.8 to 0.4 all the nodes were discovered in one trip of the reconnaissance agent. Our simulation results illustrate that the agent based node discovery technique in P2P systems performs better than flooding the network to discover peer nodes.

6 Related Work

P2P systems have become an area of active research and development since the popularity of online resource sharing services such as Freenet [4], Gnutella [6], Napster[10] and SETI@home[14]. Most of the P2P research literature [12, 15] has focussed on building scalable and distributed resource location systems for wide-area peer-to-peer applications. The Anthill framework[1] uses intelligent agents for node and resource discovery in a P2P system. Mobile agents called ants move across the nodes in a P2P network to discover resources and accomplish distributed tasks. However, most of these systems address issues related to resource discovery in P2P networks.

Techniques for intelligent node discovery in P2P networks in [16] use searches of limited diameter from the node originating the search along with data caching at every node to increase data replication. In [11] rumor mongering techniques are used to limit the number of nodes visited during node discovery and avoid flooding of the entire network by multicast messages. However, in both these techniques information about network parameters of the nodes is not incorporated in the node discovery strategy leading to an uninformed search among the peer nodes. In contrast, the technique described in this paper selects peer nodes based on an incentive function so that the nodes selected by the user for connecting to, reflect the user's interests.

7 Conclusion

In this paper, we have described an algorithm for intelligent node discovery in a P2P system using mobile agents. Our simulation results indicates that dynamic node discovery based on user incentives improves the node discovery protocol a P2P network.

This work is the first step in our research on agent enabled P2P systems. We have already started developing an incentive based model for resource discovery. We envisage that resource discovery can be improved by forming topic based sub-communities of nodes within a larger P2P network. We are also working on secure P2P interactions and studying the overhead of secure communication between agents in a P2P system.

The P2P system described in this paper is a simulation of P2P network using agent servers to model peer nodes. In the future, we plan to overlay our agent framework over a large P2P community such as JXTA [7] to compare the performance of agent enabled intelligent P2P systems with traditional P2P networks. From our preliminary results described in this paper, we envisage that agent enabled P2P systems are likely to compare favorably with network based P2P systems.

References

[1] O. Babaoglu, H. Meling and A. Montresor, "Anthill: A framework for the development of agent-based peer-to-peer systems," Proceedings of the 22nd International Conference on Distributed Computing Systems(ICDCS), Vienna, Austria, 2002, pp. 15-22.

[2] F. Cuenca-Acuna, C. Peery, R. Martin, and T. Nguyen, "PlanetP: Using gossiping to build content addressable peer-to-peer information sharing communities," Technical Report, DCS-TR-487, Department of Computer Science, Rutgers University,2002.

[3] FIPA ACL Message Structure Specification, URL http://www.fipa.org/specs/fipa00061/

[4] Freenet, URL http://www.freeproject.org

[5] P. Golle, K. Leyton-Brown I. Mironov and M. Lillibridge, "Incentives for sharing in peer-to-peer networks," Proceedings of the 2nd International Workshop on E- Commerce, Heidelberg, Germany, 2001, pp. 75-87.

[6] Gnutella, URL http://www.gnutella.com

[7] JXTA URL: http://www.jxta.org

[8] D. Lange and M. Oshima, "Programming and deploying java mobile agents with aglets," Addison Wesley, 1998.

[9] D. Miloijcic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins and Z. Xu, "Peer-to-peer computing," HP Technical Report, 2002 HPL-2002-57.

[10] Napster, URL http://www.napster.com

[11] M. Portmann and A. Seneviratne, "Cost-effective broadcast for fully decentralized peer-to-peer networks," Computer Communication, Special Issue on Ubiquitous Computing, Elsevier Science, Autumn 2002.

[12] A. Rowstron and P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems", Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms (Middleware), Heidelberg, Germany, 2001, pp. 329-350.

[13] S. Saroiu, K. Gummadi, and S. Gribble, "A measurement study of peer-to-peer file sharing systems, In MMCN, Jan. 2002.

[14] SETI URL http:// setiathome.ssl.berkeley.edu

[15] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan, "Chord: A peer-to-peer lookup service for internet applications," Proceedings of the ACM SIGCOMM Conference, San Diego, CA, USA, 2001.

[16] B. Yang and H. Garcia-Molina, "Improving search in peer-to-peer networks," Proceedings of the 22nd International Conference on Distributed Computing Systems (ICDCS'02), Vienna, Austria, July 2002, pp. 5-14.