Research of Route Algorithm Improvement in P2P Network based on Gnutella

Fan Deming, Zhuqi

Research Department, Shandong Institute of Business and Technology, YanTai, Shandong China
e-mail:fandm2008@yahoo.co.cn
e-mail:zhuqi@sdibt.edu.cn
doi:10.4156/jdcta.vol5.issue7.10

Abstract

The huge redundant messages generated by flooding which is the basic routing algorithm of unstructured P2P network, severely reduce the efficiency of the algorithm. Combining the advantages of flooding search algorithm and random walk algorithm, such as short response time, high stability, simple structure, to greatly reduce the redundancy of messages, this paper proposed an improved route search algorithm: JumpRandom-flood. The simulation results show that the algorithm can reduce the redundant messages, increase the efficiency of search, and improve the network environment while maintaining the desired coverage rate of nodes.

Keywords: Peer-to-Peer(P2P), Flooding Algorithm, Random Walk, Gnutella Network

1. Introduction

The Gnutella protocol provides a simple reliable distribution system that delivers knowledge, literature, art and music to all. It does so without need for government subsidies, corporate monopolies or buildings on the high street. There are no central servers. The network is based on peers that co-operate independently to make a library of digital content. As such, Gnutella embodies a paradigm shift from a world of corporate overlords to networked producers and consumers.

P2P has been developing rapidly which has become an important part of modern network, consisting of structured and unstructured. For structure network, the biggest difficulty is that network node is too dynamic which maintenance cost is too high. so the vast majority of commercial applications is unstructured one. Flooding algorithm is the basic algorithm in unstructured P2P network with the advantages of simple structure, easy to implement, stable to maintenance and big coverage, etc., but the drawback is also evident, resulting in too much redundant information, which increase the burden on the network, expansion of the poor. Therefore, unstructured P2P network routing algorithm has become a hot research, at the same time, for a variety of improved algorithms flood come into being.

2. Unstructured P2P Network Routing Algorithm and Its Improvement

Routing algorithm for unstructured P2P networks can be broadly classified into three categories: forwarding mechanism based on improved, based on the cache, and topology optimization.

2.1. Based on Forwarding Mechanism

The Modified BFS, Directed BFS, Intelligent BFS, and IBFS-HDI \(^{(3)}\) in Literature \([1-2]\) are part of a particular policy selecting neighbors, although they achieve the purpose of reducing redundant information, the efficiency is to be improved; \([4]\) mentioned in the random walk and expansion ring, in which random walk can reduce the number of queries an order of magnitude, but also increases the response time an order of magnitude. Extension ring is still the essence of Flooding algorithm, the difference is that this mechanisms solve the appropriate solution of TTL.
2.2. Based on The Cache

Caching methods include content caching and index cache. Content caching\(^5\), copy the contents of the file to the nodes directly, which would reduce the burden of some of the hot node as well as response time. Local Indices\(^1\), each node maintains a distance from their own no more than \(R\) nodes in the data index step, when the node receives query information, it can query for all the nodes within a radius of \(R\). A typical algorithm is LHC\(^6\), cache their neighbors resource index, send query information to the neighbors. However, in addition to save overhead, every node also send queries to neighbors, sending numbers is close to the square of node degree. The redundant information in the lower layer is still huge.

2.3. Based on Topology Optimization

Supernode structure, providing an index of local queries by supernode, the use of P2P systems is KaZaA, spanning tree is the establishment of a similar tree topology in the P2P network structure, a typical one is a LightFlood\(^7\). In this algorithm, the upper network Flooding algorithm used to reach Layer 5, each of the other node to its immediate logical neighbors to send their connectivities, nodes get all his neighbors After the maximum degree of connectivity of nodes selected as its parent node. Lookahead\(^8\) algorithm to know that the neighbor information within two hops in order to eliminate the loop. Topology optimization is significantly reduced the generation of redundant information, but the maintenance of the resource topology of the overhead caused by a problem can not be ignored, and the improvement of the network topology also bring single point of failure.

Currently, a variety of improved algorithms concentrate to reduce redundant information, but more or less introduced other problems, such as the failure of the introduction of a single point of failure by tree structure and delay caused by random walk. In view of the above problems, there comes up with a improved algorithm: random jump flooding algorithm. This algorithm tries to keep the original advantages in the flooding algorithm and reduces redundant information as well as improve query efficiency.

3. Gnutella

P2P network supported by Gnutella protocols can use small-world model to describe the overlay network topology. Small world model is an arbitrary distance between two network nodes (ie, the routing hops) is generally short, unstructured Gnutella system on a test\(^9\) showed that: In Gnutella network, more than 95% of the nodes can be searched within 7 jump; and for each node, the connection between its neighbor nodes is with high probability and cluster phenomenon is quite obvious. The algorithm and experimental simulation are based on the network in this article.

4. Random Jump Flooding Algorithm

Any topology graph can be transformed into a point for the vertex of the pyramid structure to the node requesting information for the vertices, you can become a tower Gnutella network graph, most of the nodes within the 7 layer. We can divide query information into two phases after a 7-hop: low jumps, high jump. The algorithm is divided into two parts. 1) The upper algorithm. Before Reaching the 5th layer apply jump flooding algorithm, aiming to eliminate the triangle, quadrilateral loop in two range of jump to get as many as possible of the source node while reducing redundant information when querying information in 5th layer. 2)lower algorithm. Applying the strategic Random walk to broadcast query information to prevent the redundant information at the bottom when query information.

4.1. Jump Flooding Algorithm

In jump flooding algorithm, each node saves routing information within 2 hops neighbors, namely saving routing information as well as neighbor's neighbor one. When saving the neighbor's neighbor
node routing information, first check the list of their neighbor, if the list already contains the node ID, then don’t need to save.

Query message format is divided into two types: non-forward query message and forward query information.

1) non-forward query information. Discarding query message after query, don’t forwarded.

2)forward query information. Continuing forward after inquiries and put node ID from neighbors and neighbor's neighbor in the query information list.

When a node receives a query message, first to determine whether the information receives or not, if receive, then discard, if not, query itself. Then determine the type of query information: If it is non-forward query information discard directly, if it is forward query information, forward it then.

Forwarding strategy: Remove nodes in query information attached list in neighbor nodes, then send non-forward query information to its neighbor. Remove forwarding information node and nodes from query attached list in neighbor's neighbor node list and then send forward query information to the rest neighbor’s neighbor nodes.

Figure 1, Table 1 shows, nodes A1 initiated the query information nodes. First of all neighbor nodes A1 sends a non-forward query information to B1, B2, and then add the A1 information list and neighbor’s neighbor list nodes( B1, B2, C1, C2, C3, C4) to query information, and then send forward query information to neighbor’s neighbor C1, C2, C3, C4. B1 and B2 query itself after receiving the query message, and then determine the type of information, discard the query information; C1 receives query information, find its own resources list, in the neighbor list, because there are B1, C2, D1 in the query information from A1, only need to sends the non-forward query information. In neighbor’s neighbor list, apart from sending node A1, for there is node C3 in query information attached list, eliminate it as well, only to send the forward information to node F1. C2, C3, C4 perform the same forward strategy.

The algorithm can effectively reduce triangular and quadrilateral loop within two jumps. For small-world network model, such as the Gnutella network, it can significantly reduce the large number redundancy caused by cluster phenomenon, and the higher of the node aggregation is, the more evident of the effects will be.

<table>
<thead>
<tr>
<th>node</th>
<th>neighbor</th>
<th>neighbor’s neighbor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>B1,B2</td>
<td>C1,C2,C3,C4</td>
</tr>
<tr>
<td>B1</td>
<td>A1,C1,C2,C3</td>
<td>B2,C4,D1</td>
</tr>
<tr>
<td>B2</td>
<td>A1,C1,C3,C4</td>
<td>B1,D1</td>
</tr>
<tr>
<td>C1</td>
<td>B1,C2,D1</td>
<td>A1,C3,F1</td>
</tr>
</tbody>
</table>

4.2. Random walk Routing Algorithm

As agnostic network\cite{11}, resulting in a various of redundant information when query information in lower broadcasting. According to the survey\cite{9}, with the small-world network model features 4 of redundant information generated by jumps of not more than 10%, relative to the nodes of the transmission rate is acceptable. So when the query information reach to flooding process layer 5, the switch routing algorithm with random walk. Random walk is a algorithm with small coverage of time to reduce the redundancy. Query information starts from the layer 5 nodes to use random walk, due to the fact that there are lots of source node(near four square degree of average nodes) in jump flooding.
algorithm when broadcast to the fifth network. So random walk don’t need to set high RTTL (the number of random walk hops) to reach ideal nodes coverage.

In this algorithm, when node degree is bigger than Q (Q is specified by the developer, approximately choose the average degree of network nodes), randomly selects two forwarding neighbors; if less than q, and only choose to forward a neighbor. Assuming the node average degree of q is n, starting node (i.e. layer 1 nodes) initiates inquiries, in extreme cases, the first 4-layer network is without any loops, the information forwards to layer 5 through layer 3, resulting in \((n-1)^4\) starting nodes at best. In the lower random routing algorithm, assuming that the last loop of all query information is redundant information. In this case, biggest redundant \((3/2)^{RTTL}(n-1)^4\) is received. However, at the same time, the coverage of lower node is \[\left(\frac{3}{2}\right)^1+(\frac{3}{2})^2+\cdots+(\frac{3}{2})^{RTTL-1}\] \((n-1)^4\). As RTTL becomes bigger, query efficiency (coverage of node/(coverage of node + redundancy)) is becoming bigger, the redundant information rate will be decline on the contrary.

We know from the 4.1 that when the RTTL is 4, the query message reaches the lowest level nodes, the longest path length (2 jumps flooding + 4 jumps random walk) is equivalent to the longest path flooding algorithm, which preserves the advantages of fast response in flooding algorithm. Experiments show that maintaining a certain node coverage rate greatly reduces redundancy under this condition.

4.3. Pseudo Code of Random Jump Flooding Algorithm

Assuming \(M\) is the list of the neighbors list of current node, \(N\) represents the neighbor’s neighbor list of current node, \(U\) of additional information for the query list, \(V\) controls information for the node, \(Type\) is the Boolean variable of query information the value is equal to True for non-forward information and equal False query information when forwarding; \(TTL\) initial value is 3, so check the information transmitted through twice forwarded query information to reach Level 5 nodes. \(Q\) is the specific network parameters for developers, which can select approximately the average value for the average network node degree.

Node receives a query message from node X.

If(seenMessage == true) return false; // to check if receive this information
CheckLocalMatch(); // to check local resources
\(V = U\); // to empty information in u
Type == true) return; // return if this information is non-forward
If(TTL > 1) // TTL initials is 3
{For each Y in (M-V) Type = true; // send non-forward information to its neighbors
For each Z in (N-V-X); Type = false; \(U = M+N\);
TTL--; // forward information, forward to its neighbor nodes
Return; }
If (RTTL <= 1) return false;
Else if (M.size > q) { RTTL--; // forward 2 random nodes in M}
Else { // forward one random nodes in M}
Return;

5. Simulation Experiments

Simulator uses Neurogrid, run it under Eclipse[10]. The NeuroGrid Simulator is java software designed to make it easy to simulate a number of peer to peer networks such as Freenet, Gnutella. Simulated Gnutella network with small-world characteristics. Number of nodes were selected \([6000,8000,10000,12000,14000]\) respectively, node degree \([2,10]\), the average degree of each node is 6, each node sets 2 or 3 files, each file has 3 keywords, the keywords is 1500 in network, a total of 2500 documents, \(TTL\) is 3, \(Q\) is 6.

RTTL takes 4 and 8 respectively in this simulation experiment. When RTTL is 4, the maximum path
Research of Route Algorithm Improvement in P2P Network based on Gnutella
Fan Deming, Zhuqi

length does not exceed Flooding algorithm, to preserve the advantages of fast response Flooding algorithm.

Figure 2 shows the node coverage of the two algorithms, the node coverage of flooding algorithm is [88%, 99%), indicating that the network is basically consistent with the small world model. Stochastic jump flooding algorithm coverage is lower than flooding algorithm, but even when RTTL = 4 can achieve over 50%. At the same time can be seen from Figure 3, redundant information reduces dramatically. RTTL = 4 when only flooding algorithm is [1/7, 1/6], RTTL = 8 when [1/5, 1/4]. Query efficiency is greatly increased, shown in Figure 4.

For unstructured P2P network, each node for target file is equal, this article covers the definition of query efficiency = number of nodes / total number of messages. In Figure 4, efficiency of the flooding algorithm gradually increases, with the reason that node degree doesn’t increase as the increasing of the net node. Redundancy information rate is decreasing in lower network. According to Gnutella, model information reaches to 95% coverage by 7 jumps, 14000 nodes, average degree of 6 of the network (flooding algorithm nodes only 88% coverage) has been forthcoming beyond the scope of Gnutella model.

In contrast to higher aggregation of the network, the superiority of the algorithm can be manifested. Shown in Figure 5, Figure 5 is 10000 nodes, average node degree were taken [4,6,8,10], respectively. RTTL value of 4, q change with improvement ratio (ratio = jump randomly to improve flooding query efficiency / query flooding algorithm efficiency).

Figure 6 is 10,000 nodes, RTTL were taken [4,6,8,10,15,50] of the query efficiency. when RTTL takes 50, you can think all the query information is similar to redundant information at last. Experimental results show a significant increase in RTTL won’t reduce the query efficiency, it will only have an impact on response time.
6. Conclusion

In this paper, we use the Gnutella network with the characteristic of high connectivity and low characteristic path length, combined with the advantages of the flooding algorithm and the random walk algorithm to achieve an improved routing algorithm: random jump flooding algorithm. While retaining the simple structure, stability, short response time under the premise of the node to ensure a certain coverage, greatly reduce redundant overhead. Experiments show that random jump flooding algorithm is more efficient than flooding algorithm.

7. Acknowledgment

This work is supported by Science and technology projects of Shandong Province Education Department (J10LG11).

8. References