Multicast Tree to Minimize Maximum Delay 
in Dynamic Overlay Network

Chae Y. Lee and Jin woo Baek
Dept. of Industrial Engineering, KAIST, 373-1 Kusung Dong, Taejon, Korea
{chae, bjw77}@kaist.ac.kr

Abstract
Overlay multicast technique is an effective way as an alternative to IP multicast. Traditional IP multicast is not widely deployed because of the complexity of IP multicast technology and lack of application. But overlay multicast can be easily deployed by effectively reducing complexity of network routers. Because overlay multicast resides on top of densely connected IP network,

In case of multimedia streaming service over overlay multicast tree, real-time data is sensitive to end-to-end delay. Therefore, moderate algorithm’s development to this network environment is very important.

In this paper, we are interested in minimizing maximum end-to-end delay in overlay multicast tree. The problem is formulated as a degree-bounded minimum delay spanning tree, which is a problem well-known as NP-hard. We develop tabu search heuristic with intensification and diversification strategies. Robust experimental results show that is comparable to the optimal solution and applicable in real time.

Keywords
Overlay multicast, Multicast tree, Tabu search

1. Introduction
Overlay network has become an effective way to provide advanced services in the Internet even without making any changes in the network layer. Overlay network is formed by a subset of underlying physical nodes. The connection between the overlay nodes is provided by overlay links, each of which is usually composed of one or more physical links [13]. One advanced service on overlay network is multicast.

Multicast is an important part of many next generation applications, including voice conferencing, video-on-demand and peer-to-peer file sharing. Multicast is an efficient transmission method that simultaneously provides a same type of service to a group of receivers. Router plays a significant role of replicating data packets and transmitting them to receivers, so that the source would not have to transmit the same data repeatedly. However, the deployment of multicast has not been taken by most commercial ISPs. Its deployment is difficult due to the complexity of technology and lack of applications.

Overlay multicast is highlighted as an alternative for providing multicast services in the Internet [3, 4, 5, 6, 7, 12, 17, 18]. In overlay multicast, data packets are replicated at end hosts. Each multicast member is responsible for forwarding the packets. Thus, overlay multicast does not change the network infrastructure but performs forwarding at each end host. The basic idea of multicast and overlay multicast is described in Figure 1. In Figure 1 (c), Logical tree structure is built to form an overlay multicast network regardless of the underlying physical Internet.

Most of the previous work on overlay multicast can be classified into two broad categories [16, 17]. One is end-to-end overlay and the other is proxy-based overlay. In end-to-end overlay, the end nodes are logically connected together to form an overlay network. These end nodes have the responsibility to forward data to other nodes. Proxy-based overlay is a hierarchical structure compared to end-to-end overlay. This overlay uses fixed nodes as proxies to facilitate overlay services. Most of these fixed nodes are deployed in order to achieve similar functions, such as routing and forwarding. Multicasting service is performed with the help of proxy nodes, which can duplicate data and forward them to end hosts with predefined routing algorithm. These two overlay networks are depicted in Figure 2.

In end-to-end overlay multicast, each multicast member node has the responsibility of forwarding packets. Therefore, when the multicast member nodes leave or join a multicast session, overlay multicast tree should be recovered. Such phenomenon can be explained by members logging off and disconnecting themselves from the multicast session or new members subscribing for the multicast service.

Figure 1. Overlay Multicast

In this paper, we are interested in minimizing maximum end-to-end delay in overlay multicast tree. Tabu search heuristic is developed with intensification and diversification. The rest of the paper is organized as follows. In section 2, we discuss other subjects related to our work. In section 3, we describe our problem in detail. In section 4, we present a heuristic algorithm to solve the recovery of multicast tree. In section 5, the performance of suggested algorithms is analyzed. Finally, we present our conclusion in section 6.

2. Related work
There are a number of studies on overlay multicast services in the recent literature, mostly due to the efficiency of overlay network. Some studies dealt with similar subjects to our paper.
In [17], proxy-based architecture is applied. This paper focuses on minimizing average-latency to the all end hosts. It presents a decentralized scheme that organizes the Multicast Service Nodes (MSNs) into an appropriate overlay structure that is particularly beneficial for real-time applications. However, our paper tries to minimize maximum overlay delay. As a measure, minimizing the maximum delay is fairer than minimizing the average delay. We formulated this problem and suggested a tabu search algorithm.

In [15], to cope with the node failure in the overlay multicast network, the employment of MSNs was considered and it allowed relatively high processing performance to cover the disconnected nodes. This paper only presents a centralized algorithm with Lagrangean relaxation as its solution tool. Centralized algorithm, however, is not an effective method for rapid recovery of the network. For immediate and efficient recovery, we introduced an instant strategy and a centralized strategy at the same time. Our recovery strategy is more useful and realistic.

3. IP Formulation for Overlay Multicast Tree

Overlay multicast tree can be evaluated with the quality of data delivery path, the robustness of overlay multicast and control overhead at members [13]. The quality of the tree is measured with metrics such as node stress and node degrees. Since end nodes are potentially less stable than routers, for robust overlay multicast tree, it is important to alleviate the effect of member failures. Rapid recovery of the tree and packet delivery to other members are necessary. Finally, for scalability of the overlay multicast to large member groups control overhead at members should be low.

In this paper, we are interested in the robustness of overlay multicast. For the robustness two critical points are considered in the process of recovery. We want to provide multicast service without any stoppage. Therefore, the tree needs to be rapidly recovered when failure occurs. Moreover, we want multicast services with minimum delay. To satisfy these two conditions, instant and centralized recovery strategy of overlay multicast tree is introduced.

Instant strategy is an immediate solution procedure, which rapidly recovers the tree for continuous services. This strategy can be applied whenever a node leaves or joins the multicast group and causes failure in multicast tree. The strategy suggests that disconnected nodes should be attached to end nodes where delay would be the least. However, instant strategy does not guarantee overlay multicast tree with the minimum delay.

To obtain better overlay multicast tree with minimum delay centralized strategy is considered. The strategy periodically reconstructs the overlay multicast tree. Specifically, using the centralized strategy, we want to find an overlay multicast tree that minimizes maximum overlay delay for all multicast members. To guarantee the quality of data delivery path packet processing ability of each member node is restricted by degree bound. Therefore, this problem is formulated into degree-bounded minimum delay spanning tree.

Consider a graph $G = (V, E)$ where $V$ is the set of multicast nodes and a source and $E$ is the set of links as a full-meshed overlay network. A node $m \in V$ represents a multicast member that wants to receive services from the source. If no path exists between the source and the multicast member $m$, the tree recovery is impossible. By assuming at least one path between the source and node $m$, the failed multicast tree is recovered to connect every multicast member.

Let $x^m_{ij}$ be a binary variable to represent a link $(i, j)$ that is employed for a path between the source and multicast member $m$. If link $(i, j)$ is employed to connect the source and node $m$, then $x^m_{ij} = 1$. Otherwise, $x^m_{ij} = 0$. Then the following relations hold for every node including the source $s$.

$$\sum_{j \in J} x^m_{ij} - \sum_{j \in J} x^m_{ji} = 1, \quad \text{for all } m \in V \text{ and } i = s$$

$$\sum_{j \in J} x^m_{ij} - \sum_{j \in J} x^m_{ji} = 0, \quad \text{for all } m \in V \text{ and } i \notin \{s, m\}$$

$$\sum_{j \in J} x^m_{ij} - \sum_{j \in J} x^m_{ji} = -1, \quad \text{for all } m \in V \text{ and } i = m$$

Let $y_j$ be a binary variable to represent the adoption of link $(i, j)$ for paths in the overlay multicast tree, then the following equation holds.

$$x^m_{ij} \leq y_j, \quad \text{for all } m \in V \text{ and } i \neq j$$

The above constraint represents that link $(i, j)$ has to be selected in a tree to create the path between the source and multicast member $m$.

Now, we consider the degree constraint of a multicast member node. By letting $w_i$ be the degree constraint of node $i$, we have

$$\sum_{j \in J} y_j + \sum_{j \in J} y_j \leq w_i, \quad \text{for all } i \in V$$

To build up a spanning tree, we have

$$\sum_{j \in J} y_j = n - 1, \quad \text{for all } i \in V$$

Our objective in the multicast tree recovery is to minimize maximum overlay delay for all end hosts. By letting $d_{ij}$ be the link delay and $Z$ be the maximum end-to-end delay, we have the following equation to minimize $Z$.

$$\sum_{(i,j) \in E} d_{ij} x^m_{ij} \leq Z, \quad \text{for all } m \in V$$

From the above discussion, we have the following binary integer programming formulation.

Minimize $Z$

subject to

$$\sum_{(i,j) \in E} d_{ij} x^m_{ij} \leq Z, \quad \text{for all } m \in V$$

$$\sum_{j \in J} x^m_{ij} - \sum_{j \in J} x^m_{ji} = 1, \quad \text{for all } m \in V \text{ and } i = s$$

$$\sum_{j \in J} x^m_{ij} - \sum_{j \in J} x^m_{ji} = 0, \quad \text{for all } m \in V \text{ and } i \notin \{s, m\}$$

$$\sum_{j \in J} x^m_{ij} - \sum_{j \in J} x^m_{ji} = -1, \quad \text{for all } m \in V \text{ and } i = m$$

$$x^m_{ij} \leq y_j, \quad \text{for all } m \in V \text{ and } i \neq j$$

$$\sum_{j \in J} y_j + \sum_{j \in J} y_j \leq w_i, \quad \text{for all } i \in V$$
4. Tabu search

Tabu search has achieved widespread successes in solving practical optimization problems. Applications are rapidly growing in areas such as resource management, process design, logistics, technology planning and general combinatorial optimization [2]. A tabu search incorporates three general components [1]. 1) Short-term and long-term memory structure, 2) Tabu restrictions and aspiration criteria, and 3) Intensification and diversification strategies.

Intensification strategies utilize short-term memory function to integrate features or environments of good solutions as a basis for generating still better solutions. Such strategies focus on aggressively searching for a best solution within a strategically restricted region. Diversification strategies, as their name suggests, are employed in long-term memory function, redirect the search to unvisited regions of the solution space.

4.1 Initial Multicast Tree

For the overlay multicast services, the member nodes organize themselves into an initial data delivery tree. We suggest two algorithms for initialization: greedy algorithm and path attachment algorithm. We assume that the node-to-node delay is known by the source node. Each multicast member node sends hello message to the source to receive packets. This hello message includes information about node degree and delay. Thus, the source has the knowledge about all the member nodes.

In greedy algorithm the procedure starts by selecting a node with least delay from the source node. The degree of source is reduced by one. From either the source or any of the connected nodes, the procedure selects a node that can be connected with the least link delay. If the degree of any of the connected nodes reaches zero, that node is not considered in the process. This process is continued until all the nodes are connected.

Path attachment algorithm applies the degree bound after the shortest path to each node is obtained. The algorithm first finds shortest path from source to each node by Dijkstra algorithm [19]. The procedure selects the longest path and attaches it to the source. Then the procedure selects the second longest path and connects the path to the node that satisfies the degree bound with minimum delay. This process is continued until all nodes are connected. Figure 3 shows initial trees by two algorithms with overlay multicast network given in (a).

4.2 Intensification with short-term memory

Intensification process utilizes short-term memory to integrate features of good solutions. Such strategies focus on aggressively searching for a best solution within a strategically restricted region. A move remains tabu during a certain period (or tabu time size) to help aggressive search for better solution.

Since our objective in the dynamic overlay network is to minimize the maximum delay, at each iteration we select a node with maximum delay for a move. We consider two types of moves: “node swap” and “node reconnection”. In node reconnection a target node is reconnected to another node that satisfies the degree bound with least delay. This procedure is repeated until the solution is not improved. An end node of a path with maximum end-to-end delay is selected as the target node. When this target node does not improve the solution, the parent of the end node is considered as another target node. The parent and all its children nodes are moved in the reconnection.

Note that the link delay from the reconnected node to the parent node can be improved even though the delay from the parent to its children is not changing. Nevertheless, if the solution is not improved, the procedure reconnects the end node to its grand parent node, and parent node to other node that satisfies the degree bound with least delay. Solution will be updated when it is improved. Otherwise, reconnection procedure is terminated. After the reconnection, node swap move is started. Node swap is implemented by swapping the target node with a node with least delay. This swap move thus allows an uphill move in the tabu search for better solution. After this swap move, reconnection and swap is repeated.

4.3 Diversification with long-term memory

Diversification strategies, as their name suggests, are designed to drive the search into new regions. In general, the performance of tabu search becomes significantly stronger by including long term memory and its associated strategies. In this study frequency-based memory is used for diversification. Links that are not frequently in the previous multicast trees are chosen for a new tree. To restart the search after each pass, the procedure excludes n-1 most frequently used links during the pass. Rest of the links is considered to have an initial tree for diversification.

4.4 Overall tabu search procedure
Based on the intensification and diversification strategies in the previous sections, we describe the overall procedure of the tabu search. First, we generate initial overlay multicast tree. Starting from the initial tree move operations are implemented. Whenever a move is implemented, we update the current solution $S_{current}$ and the best solution $S_{best}$. Tabu list and the frequency of each link in the solution $F(i,j)$ are also updated. The intensification process is continued as far as the best solution is improved. When moves result in no improvement consecutively for $N_{max}$ iterations the diversification process is performed with a new overlay multicast tree. When the number of diversification reaches $D_{max}$, the overall procedure is terminated. The overall tabu search procedure is explained as follows.

Step1 Initial tree generation
Obtain initial overlay multicast tree by algorithms in Section 4.1

Step2 Initial parameters of tabu search
Set $N=0$, $D=0$ and $F(i,j)=0$
Tabu list is empty

Step3 Intensification
Select a target to apply move
Apply reconnection moves until the solution is not improved
Apply swap move
Update $S_{current}$
Update $S_{best}$, tabu list and $F(i,j)$
If $N < N_{max}$, repeat this step
If $N > N_{max}$ and $D < D_{max}$, go to step 4
If $D > D_{max}$, stop the tabu search

Step4 Diversification
Obtain a new tree using $F(i,j)$
Restart the tabu search with the initial tree
Set $N=0$ and $F(i,j)=0$
Go to Step 3

5. Computational Results

In this section, we discuss the computational results of the tabu search for the degree-bounded minimum delay spanning tree. Four different sizes of overlay multicast networks are generated each with 10, 30, 50 and 100 nodes. The node degree is assumed to be ranged 2 ~ 4. In each network, links are randomly generated such that the delay of each overlay link is ranged 1 ~ 10 by assuming maximum 10 physical hops. The algorithm for Tabu search is implemented in C language and run on a 2.40 GHz Intel Pentium 4 based on personal computer with 512 Mbyte of memory under Windows XP.

We first test two initial solution strategies: greedy algorithm and path attachment algorithm. Ten problems with 50 nodes are tested as in Table 1. As shown in the table, path attachment algorithm is superior to the greedy algorithm. We thus start the tabu search with the initial solution by the path attachment algorithm.

Before testing the performance of our tabu search, we need to tune the tabu parameters: tabu list size, $N_{max}$ for the intensification procedure and $D_{max}$ of the diversification. Tabu list size represents the number of iterations during which a target node is forbidden to be adopted in move operation. By assuming that an appropriate tabu list size is proportional to the number of nodes, we perform tests with 50 nodes. Figure 4 shows that $0.2N$ is suitable for tabu list size with 50 nodes.

Test for $N_{max}$ is performed as in Figure 5. The figure shows that appropriate value for $N_{max}$ is 0.6N for 10 nodes, 0.6N for 30 nodes, 0.8N for 50 nodes and 0.8N for 100 nodes. The test on $D_{max}$ is also performed as in Figure 6. By testing ten problems the portion that gives no further improvement for the successive diversification is plotted in the figure. From the figure, it seems to be reasonable to perform two diversifications for problems with 10 nodes, three for 30 nodes, six for 50 nodes and nine for 100 nodes.

Computational results of the proposed tabu search are shown in Table 2,3,4,5. CPLEX [14] is employed to compare the solutions. Due to the exponential growth of the branches in the process of CPLEX, it fails to obtain the optimal solution even with 10,000 seconds running time for networks with 30, 50 and 100 nodes. Especially, in the network with 100 nodes CPLEX experiences memory problem. Table 2 shows the results of tabu search with 10 nodes. The proposed tabu search generated optimal solutions in all cases with 10 nodes. Table 3 and 4 show the results of tabu search with 30 and 50 nodes respectively. The proposed tabu search generated optimal solutions in two cases with 30 and 50 nodes. The gap in worst case reaches 15.4% in problems with 30 and 50 nodes. Table 5 shows the results of tabu search with 100 nodes. CPLEX is out of memory in the problems with 100 nodes. Thus, we fail to obtain the near optimal solution. Computational experiments are performed in overlay multicast network 10, 30, 50 and 100 nodes. The efficiency of tabu search in CPU seconds is illustrated as the number of multicast nodes increases.

The effectiveness of diversification in this tabu search is demonstrated in Figure 7. The figure shows results of tabu search with and without diversification. With diversification, the maximum end-to-end delay of the overlay multicast tree is decreased approximately by 1.5% for 10 nodes, 2.9% for 30 nodes, 3.7% for 50 nodes and 3.5% for 100 nodes.

We have additionally carried out tests in order to observe the effect of node degree bound. Figure 8 shows the effect of no degree. To analyze the effect of node degree, the degree of all nodes is fixed to 2, 3 and 4 respectively except for the source node. As shown in figure 8, the decrease of average maximum delay seems to be sensitive to the node degree in problems with 10, 30, 50 and 100 nodes.

<table>
<thead>
<tr>
<th>Table 1. Test of Feasible Initial Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>Average</td>
</tr>
</tbody>
</table>
### Table 2. Computational results of tabu search with 10 nodes

<table>
<thead>
<tr>
<th>Problem</th>
<th>Tabu Search Solution (TS)</th>
<th>CPU seconds</th>
<th>CPLEX Solution (OPT)</th>
<th>CPU seconds</th>
<th>GAP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>1.11</td>
<td>8</td>
<td>5.55</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>1.01</td>
<td>6</td>
<td>0.34</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>1.12</td>
<td>6</td>
<td>0.17</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1.10</td>
<td>8</td>
<td>0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>1.13</td>
<td>10</td>
<td>12.38</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>1.12</td>
<td>9</td>
<td>0.2</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>1.03</td>
<td>9</td>
<td>7.56</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>1.05</td>
<td>6</td>
<td>0.34</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>1.12</td>
<td>7</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>1.09</td>
<td>7</td>
<td>0.17</td>
<td>0.00</td>
</tr>
</tbody>
</table>

* GAP: (TS-OPT)/OPT

### Table 3. Computational results of tabu search with 30 nodes

<table>
<thead>
<tr>
<th>Problem</th>
<th>Tabu Search Solution (TS)</th>
<th>CPU seconds</th>
<th>CPLEX Solution (OPT)</th>
<th>CPU seconds</th>
<th>GAP**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>60.08</td>
<td>15</td>
<td>10,000*</td>
<td>0.067</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>62.94</td>
<td>17</td>
<td>10,000*</td>
<td>0.059</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>60.39</td>
<td>17</td>
<td>10,000*</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>60.56</td>
<td>13</td>
<td>10,000*</td>
<td>0.154</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>60.25</td>
<td>15</td>
<td>10,000*</td>
<td>0.067</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>62.78</td>
<td>17</td>
<td>10,000*</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>63.23</td>
<td>16</td>
<td>10,000*</td>
<td>0.063</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>61.86</td>
<td>17</td>
<td>10,000*</td>
<td>0.059</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>62.00</td>
<td>15</td>
<td>10,000*</td>
<td>0.067</td>
</tr>
<tr>
<td>10</td>
<td>18</td>
<td>60.39</td>
<td>16</td>
<td>10,000*</td>
<td>0.125</td>
</tr>
</tbody>
</table>

*: Terminated by the time limit, ** GAP: (TS-OPT)/OPT

### Table 4. Computational results of tabu search with 50 nodes

<table>
<thead>
<tr>
<th>Problem</th>
<th>Tabu Search Solution (TS)</th>
<th>CPU seconds</th>
<th>CPLEX Solution (OPT)</th>
<th>CPU seconds</th>
<th>GAP**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>73.76</td>
<td>27</td>
<td>10,000*</td>
<td>0.074</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>73.44</td>
<td>24</td>
<td>10,000*</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>74.19</td>
<td>23</td>
<td>10,000*</td>
<td>0.130</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>74.12</td>
<td>28</td>
<td>10,000*</td>
<td>0.036</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>74.60</td>
<td>22</td>
<td>10,000*</td>
<td>0.045</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>73.48</td>
<td>20</td>
<td>10,000*</td>
<td>0.100</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>74.05</td>
<td>26</td>
<td>10,000*</td>
<td>0.154</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>73.01</td>
<td>22</td>
<td>10,000*</td>
<td>0.091</td>
</tr>
<tr>
<td>9</td>
<td>24</td>
<td>73.36</td>
<td>22</td>
<td>10,000*</td>
<td>0.091</td>
</tr>
<tr>
<td>10</td>
<td>24</td>
<td>74.09</td>
<td>24</td>
<td>10,000*</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*: Terminated by the time limit, ** GAP: (TS-OPT)/OPT
6. Conclusion

We have presented two efficient recovery strategies of multicast tree in the end-to-end overlay network. An instant strategy was introduced to provide rapid recovery and a centralized strategy was proposed to achieve performance close to the optimal level. Overlay multicast tree was represented as a degree-bounded minimum maximum-delay spanning tree, which is a problem known as NP-hard. Thus, a tabu search was introduced as part of the centralized strategy. We compared the results of the suggested tabu search with that of the IP formulation with several simulation studies. As the result, the proposed tabu search and the strategy as a whole showed excellent results, where the solution was very close to the optimal, but derived much more rapidly.

References


