Branch-and-Cut Algorithm for the Clustering Problem in ATM VP-based Leased Line Network Design

Deokseong Kim, Kyungsik Lee, Sungsoo Park, Kyungchul Park

1. Introduction

In this paper, we consider a clustering problem for ATM VP-based leased line network design. We are given a set of access nodes with demand requirements, a set of candidate edge nodes, and a set of edge facilities. The problem is to determine a partitioning of the network into clusters, and the facility assignment of access nodes to the edge nodes. The problem is to find these decisions with minimum cost such that the demand of all the access nodes should be satisfied without violating the capacity constraints of the edge facilities. The cost is the sum of facility installation costs and the demand flow costs between access nodes, which depend on the demand of the access nodes and unit flow costs between them.

Generally, the clustering problem of the graph into K clusters with minimizing sum of costs for the edges inside the cluster variants of this clustering problem are of interest in the context of this paper. When constructing these clusters, we may devise some valid inequalities; see Barah al. [1], Conforti et al. [2, 3], Johnso et al. [4, 5], and many others. The clustering problem has also been studied for the telecommunication network design problem. For example, Laguna [7] have studied clustering problems for the design of the SONET rings. Park et al. [10] have also studied a clustering problem with node compatibility and new survivability requirements.

In this paper, we propose an integer programming model and devise some valid inequalities for the inequalities, we propose a cutting plane algorithm. The algorithm has provided the optimal solutions in reasonable computational time. The remainder of the paper is organized as follows. In the next section, we present the formulation of the problem. In section 3, we analyze the clustering problem. In section 4, we give a description of the overall algorithm. In section 5, computational results are given. Finally, in section 6, the conclusion is presented.

2. Formulation

In this section, we present the formulation of the clustering problem. For the modeling of the problem, we assume that the problem is considered...
on the physical graph $G=(N,E)$. Following notations and decision variables to be used modeling.

**Notations**

$N$: Set of access nodes

$H$: Set of candidate edge nodes, $H \subseteq N$.

$H(i)$: Set of candidate edge nodes that compatible with access node $i$.

$K$: Set of candidate edge facilities

$G=(N,E)$: Adjacency graph, where $(i, j) \in E$ if only if the access nodes $i$ and $j$ are adjacently.

$d_i$: $(\text{Symmetric})$ Demands in the number of $u$ demand types between access nodes $i$ and $i \neq j \in N$.

$D$: Total demands from to access node $i$, $i \in N$.

$B$: Capacity of edge facility of type $k$, $k \in H$.

$W$: Cost of edge facility of type $k$, $k \in K$.

$u_i$: Unit flow cost between nodes $i$ and $j$, $i \neq j \in N$.

**Decision variables**

$x_{ih}$: 1 if edge facility of type $k$ is installle edge node $h$, and 0 otherwise, where $h \in H$, $k \in K$.

$y_{ij}$: 1 if access node $i$ is connected to edge node $j$, and 0 otherwise, where $i \neq j \in N$, $h \in H$.

$z_{ih}$: 1 if access node $i$ is connected to edge node $j$ and access node $j$ is connected to edge node $g$, and 0 otherwise, where $i \neq j \in N$, $h$, $g \in H$.

In the notations, we assume that $B_1 < B_2 < \cdots < B_{|K|}$ and $W_1 < W_2 < \cdots < W_{|H|}$ and $|D| = \sum d_i$ for $i \in N$. Then, the mathematical model of the clustering problem (CLG) can be formulated as follows.

**Formulation (CLG)**

\[
\min \sum_{k \in H} \sum D_{ijk} x_{ih} + \sum_{k \in K} y_{ij} y_{jk} + \sum_{k \in K} y_{ij} z_{ig} y_{gh} \quad \forall i \in N \tag{1}
\]

\[
\text{s.t.} \quad \sum_{j \in H} y_{ij} = 1 \quad \forall i \in N \tag{2}
\]

\[
\sum_{i \in N} x_{ih} = y_{ih} \quad \forall h \in H \tag{3}
\]

\[
y_{ij} + y_{jk} - y_{ij} y_{jk} \leq 1 \quad \forall i \neq j \neq k, h \in H \tag{4}
\]

\[
y_{ij} \leq y_{ij} \quad \forall i \neq j, h \in H \tag{5}
\]

\[
\sum_{i \in H} y_{ij} \geq y_{ij} \quad \forall i \neq j, h \in H \tag{6}
\]

\[
y_{ih} \leq B_i \quad \forall h \in H, k \in K \tag{7}
\]

\[
y_{ij} \leq y_{ij} \quad \forall i \neq j, h \in H, k \in K \tag{8}
\]

The objective function represents the sum of the facility installation costs, the total intra-cluster traffic flow cost, and the total inter-cluster traffic flow cost. Constraints (1) provide the demand restriction for several hub facility types on each cluster. Constraints (2) ensure each node to be in some cluster, (3) ensure that the hub facility should be installed if the access node $h$ is selected as hub node, and (4) and (5) represent the equation $z_{ihj} = y_{ih} y_{ij}$. Constraints (6) means edge connectivity constraints on $G$.

We mention that CLG is a very hard integer programming problem and it can be proved that even the feasibility checking problem associated to it is NP-complete. It can be shown by transforming Quadratic Assignment Problem into CLG (Garey and Johnson [4]).

### 3. Analysis of the Clustering Problem

As mentioned in the previous section, CLG is a very hard problem. In this section, we investigate the polyhedral structure of the problem and derive some useful results that will be used in devising a branch-and-cut algorithm for the problem.

#### 3.1. Preprocessing and Tightening the Constraints

We define demand function $f_k(S) = \sum_{i \in E(H)} D_i + \sum_{i \in E(H)} \sum_{j \in E(H)} d_{ij} n_{ij}$ for $h \in H$. Then this demand function $f_k(S)$ is nondecreasing for $S \subseteq T \subseteq N$. Therefore we have following proposition.

**Proposition 1.** Following inequality is valid

\[
\sum_{h \in H} x_{ih} \geq \left\lceil \max_{h \in H} f_k(N) - B_k \right\rceil \quad \forall i \in N \tag{9}
\]

(proof) refer to Kim et al. [6]

**Proposition 2.** Following knapsack constraints imply the constraints (1).

\[
\sum_{h \in H} y_{ih} \leq x_{ih} \quad \forall h \in H \tag{10}
\]

(proof) refer to Kim et al. [6]

**Proposition 3.** Following constraints imply the constraints (4) and (5).

\[
y_{ij} - \sum_{i \neq h \in H, k \in K} \alpha_{ijh} = 0 \quad \forall i \neq j \in N, h \in H, k \in K \tag{11}
\]

The validity of constraints (11) can be referred from Mauricio et al. [8]. Note that the constraints (11) can replace the logical constraints (4) and (5), and the model using constraints (11) gives at least as good lower bound as the model using the constraints (4) and (5) in case that constraints (7) and (8) are relaxed.

#### 3.2. Development of Valid Inequalities

In this section, we consider a subproblem, which is a clustering problem corresponding to the edge node $h$ in the original clustering problem. Let
In this polytope, we assume that all the access nodes are compatible with each other and can be assigned to edge node $h$. If the access node $i$ is not compatible with edge node $h$, the node $i$ can be made to be compatible with edge node $h$ by setting the flow cost $u_{ih}$ to be sufficiently large cost. Also, we assume, for simplicity, that for each $s \in \mathbb{N}$ such that $|s| \leq s$ satisfies $\sum B_i \leq B_j$.

Now, we will introduce some valid inequalities for $P_s(N)$. First, we consider the equations which hold for all solutions in $P_s(N)$ to obtain the dimension of $P_s(N)$. Let a complete graph $H = (N, F)$. Then, following equations hold for all solutions in $P_s(N)$.

$$\begin{align*}
\sum_{i \in N} x_{ih} &= y_{ih} \\
\sum_{j \in F} z_{ijk} &= y_{jk} \quad \text{for all } (h, j) \in F
\end{align*}$$

Using these equations, we have following proposition

**Proposition 4.** $\dim(P_s(N)) = |V|(|N| - 1)/2 + |K|.$

proof) refer to Kim et al. [6]

**Definition 1.** For $C \subseteq N$, $C$ is an independent set for capacity $B_i$ if $\sum D_{ii} \leq B_i$; otherwise $C$ is a independent set for capacity $B_i$ if $B_i < \sum D_{ii} \leq B_i$.

**Theorem 5.** For a given $C \subseteq N$, suppose $C$ is a minimal dependent set for capacity $B_i$. Then, the following inequality

$$\sum_{k \in N} x_{ik} + |C| \sum_{k \in C} x_{ik} \leq \sum_{k \in N} y_{ik} + |C| \sum_{k \in C} y_{ik} \quad (14)$$

is valid for $P_s(C)$

proof) refer to Kim et al. [6]

**Theorem 6.** Let $C$ be a dependent set for capacity $B_i$, which includes edge node $h$. Let $T$ be a tree of $H$ that form a spanning tree on the nodes in $C$. Let $d_i$ be the degree of the node for each $i \in C$ in tree $T$. In case that $h$ is a leaf node, the following inequality

$$\sum_{k \in N} z_{ikh} + \sum_{k \in C} x_{ik} \leq \sum_{k \in N} y_{ik} + y_{ih} \quad (15)$$

defines a facet of $P_s(C)$ and only if $C \setminus \{i\}$ is an independent set for capacity $B_i$ for every leaf node $i$ of the tree induced by $T$.

4. Branch-and-cut Phase

The branch-and-cut procedure is very similar to the branch-and-bound procedure except that we solve the problem at each node in the enumeration tree by using the strong cutting planes. That is, we perform the same tasks as in the node 0 in the enumeration tree except for the fact some variables are set to 1 or 0 in the linear program. For the details of the branch-and-cut algorithm, see Padberg and Rinaldi [9].

Note that a given solution $(y', x')$ is integral if and only if $y'$ is integral. Therefore, we only set $y_{ih}$, $i \in N, h \in H$ variables and $x_{ik}$, $h \in H, k \in K$ variables to be binary in the branch-and-cut phase. When we get a fractional optimal solution $(x', y', z')$ at a node of the enumeration tree, we first select a variable $x_{ik}$ among the fractional variables, where $x_{ik}$ is the most fractional variable among the $x'$ variables. When $x'$ variables are all integral, we proceed the same branching procedure for the $y'$ variables. After the variable selection, we generate two new sub nodes with the bound adjusted. The best bound rule is used for the branching node selection in the enumeration tree.

5. Computational Results

We have tested the proposed algorithms on some randomly generated problem instances and the real problem instances. These test classes are generated according to the sizes of the edge nodes and access nodes. For each randomly generated problem class, we generated 10 problem instances. The access nodes are generated in the Euclidean plane of size of $[0, 50]$ by $[0, 50]$. The demands of access nodes and edge nodes are set to be Euclidean distances between the access nodes and edge nodes.

We set the capacities of the edge facilities as integer multiples of STM1 such as 22 STM1, 45 STM1, 716 STM1, etc., which are based on the real facility capacities. Then, we set the cost parameter of edge facility to be 10000 for the test of the randomly generated data instances, referring
to the cost ratio of real switching systems and traffic flow costs.

For an LP solver and constraints addition routine, we use the CPLEX 6.0 callable mixed integer library. All tests are performed on a Pentium PC (333MHz).

Table 1 shows the computational results of the proposed algorithm. The column #CUT refers to the number of the cuts generated. The column Gap0(%) refers to the gap between the optimal LP solution and optimal integer solution without any cut. The column Gap1(%) refers to the gap between the optimal LP solution and optimal integer solution with cut. The column #Node refers to the number of nodes generated in the branch-and-bound stage. Finally, the column Time refers to the execution time in second needed to solve the problem.

The results show that Gap0 has been reduced to Gap1 considerably after applying the cutting plane. Moreover, several test data of the small sizes are solved to optimal solutions without invoking the branch-and-bound phases. Therefore the results show that our algorithm gives the integral optimum solution in a reasonable time.

6. Conclusion

In this paper, we considered the clustering problem in ATM VP-based leased line network design. To solve this problem, we developed some valid inequalities and applied the branch-and-cut algorithm. To strength the LP relaxation of the problem, we also developed the preprocessing procedure. Computational results show that the proposed algorithm can provide the optimal solutions in reasonable time and the algorithm could also give the final LP relaxation with strong lower bound using the cutting planes.

The proposed algorithm can be applied to solve the clustering problems such as clique partitioning problem and graph partitioning problem, or can be used as a design module to solve the telecommunication network design problem that contains backbone and local networks.

References


<table>
<thead>
<tr>
<th>#CUT</th>
<th>Gap0</th>
<th>Gap1</th>
<th>#Node</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td>5.1</td>
<td>3.30</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>(3, 9)</td>
<td>Mx</td>
<td>25.0</td>
<td>5.19</td>
<td>0.00</td>
</tr>
<tr>
<td>Mn</td>
<td>1.0</td>
<td>0.60</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Avg</td>
<td>15.1</td>
<td>1.90</td>
<td>0.01</td>
<td>0.2</td>
</tr>
<tr>
<td>(3, 17)</td>
<td>Mx</td>
<td>36.0</td>
<td>5.06</td>
<td>0.08</td>
</tr>
<tr>
<td>Mn</td>
<td>3.0</td>
<td>0.25</td>
<td>0.00</td>
<td>0.44</td>
</tr>
<tr>
<td>Avg</td>
<td>22.4</td>
<td>3.04</td>
<td>0.04</td>
<td>0.8</td>
</tr>
<tr>
<td>(5, 17)</td>
<td>Mx</td>
<td>46.0</td>
<td>4.71</td>
<td>0.29</td>
</tr>
<tr>
<td>Mn</td>
<td>6.0</td>
<td>1.60</td>
<td>0.00</td>
<td>3.02</td>
</tr>
<tr>
<td>Avg</td>
<td>32.3</td>
<td>2.42</td>
<td>0.02</td>
<td>12.37</td>
</tr>
<tr>
<td>(4, 25)</td>
<td>Mx</td>
<td>62.0</td>
<td>4.23</td>
<td>0.12</td>
</tr>
<tr>
<td>Mn</td>
<td>9.0</td>
<td>0.23</td>
<td>0.00</td>
<td>3.46</td>
</tr>
<tr>
<td>Avg</td>
<td>58.2</td>
<td>2.84</td>
<td>0.29</td>
<td>143.63</td>
</tr>
<tr>
<td>(16, 25)</td>
<td>Mx</td>
<td>159.0</td>
<td>3.94</td>
<td>1.24</td>
</tr>
<tr>
<td>Mn</td>
<td>14.0</td>
<td>1.38</td>
<td>0.00</td>
<td>8.74</td>
</tr>
<tr>
<td>Avg</td>
<td>85.0</td>
<td>2.26</td>
<td>0.17</td>
<td>6.4</td>
</tr>
<tr>
<td>(5, 32)</td>
<td>Mx</td>
<td>243.0</td>
<td>3.27</td>
<td>0.80</td>
</tr>
<tr>
<td>Mn</td>
<td>8.0</td>
<td>1.01</td>
<td>0.00</td>
<td>27.24</td>
</tr>
<tr>
<td>Avg</td>
<td>86.9</td>
<td>3.02</td>
<td>0.24</td>
<td>12.4</td>
</tr>
<tr>
<td>(7, 32)</td>
<td>Mx</td>
<td>193.0</td>
<td>4.05</td>
<td>0.79</td>
</tr>
<tr>
<td>Mn</td>
<td>20.0</td>
<td>2.18</td>
<td>0.00</td>
<td>78.49</td>
</tr>
</tbody>
</table>
