New Formulations for the Generalized Traveling Salesman Problem

IMDAT KARA
Baskent University, Department of Industrial Engineering, Baglica Campus, Ankara,TURKEY
ikara@baskent.edu.tr

HUSEYIN GUDEN
Baskent University, Department of Industrial Engineering, Baglica Campus, Ankara,TURKEY
hsyngdn@baskent.edu.tr

OZGE N. KOC
Baskent University, Department of Industrial Engineering, Baglica Campus, Ankara,TURKEY
ozgekoc@hotmail.com

***Abstract*:** The Generalized Traveling Salesman Problem (GTSP) is an extension of the well known Traveling Salesman Problem (TSP). The GTSP is defined on a graph in which the nodes (customers or vertices) are grouped into a given number of clusters (node sets). Solution procedures for the GTSP are generally focused on transforming the problem to the TSP and applying the exact or heuristic solution methods developed for the TSP. There exist a few integer programming formulations for the GTSP some of which are exponential size with respect to number of the nodes. In this paper, we propose two new formulations for the GTSP with polynomial size with respect to number of the nodes. For preliminary computational analysis, GTSP instances from TSPLIB are solved by proposed formulations and also by the previously existing formulations in the literature. Performances of the formulations in terms of linear programming relaxations and CPU times are analyzed. We observe that, performances of the proposed formulations are better than the existing formulations in terms of these two evaluation criteria.

***Key-Words*:** Generalized Traveling Salesman Problem, Mathematical Modeling, Integer Programming, Logistics, Routing

1 Introduction

The TSP, the problem of finding the minimum cost Hamiltonian Tour on a given graph, is one of the most studied combinatorial optimization problems. There exist a large number of application areas, and exact or heuristic solution methods for the TSP (Lawler et al. 1984, Gutin and Punnen 2002, Dinh and Mamun, 2004; Aziz et al., 2009). The Generalized Traveling Salesman Problem (GTSP) is an extension of the TSP where the nodes of the graph is grouped into clusters and the problem is to find a minimum cost Hamiltonian Tour which visits each cluster once and visits (exactly or at least) one node from each cluster. The GTSP is a special case of the generalized network design problem (Pop et al., 2007). The GTSP defined as "exactly one node from each cluster" or "at least one node from each cluster" according to the number of nodes must be visited from each cluster. Laporte and Nobert (1983) show that, "at least one node from each cluster" coincide with "exactly one node from each cluster" when the distance matrix is Euclidean. In this paper, we consider the "exactly one node from each cluster" case. In accordance with Laporte and Nobert (1983), the GTSP is defined by Henry and Lapordere in 1969, Srivastava et al. in 1969 and Saksena in 1970 independently and dynamic programming is proposed as a solution procedure.

The GTSP and its variants may arise in real-life applications such as loop material flow design, post-box collection, arc routing, computer operations, manufacturing logistics, distributions of goods by sea to the potential number of harbors (see Laporte et al. 1996, Laporte and Palekar 2002, Ben-Arieh et al. 2003).

The GTSP is NP-hard since it reduces to the TSP when each cluster contains only one node. We may group the solution approaches of the GTSP under the following headings:

i) *Special solution procedures based on a decision model* (Laporte and Nobert, 1983; Laporte, Mercure and Nobert, 1987; Noon and Bean, 1991; Fishetti, Gonzales and Toth, 1995, 1997, 2002) and/or *to use a decision model directly* (Kara et. al., 2005; Kara and Demir, 2006; Pop, 2007),

ii) *Solving the GTSP by transforming it to TSP and solving related TSP.* (Noon and Bean, 1993; Lien-Ma-Wah, 1993; Dimitrijevic and Saric, 1997; Laporte and Semet, 1999; Ben-Arieh et al., 2003) and,

iii) *Special Heuristics* (Renaud and Boctor, 1998; Huang et al. 2005; Snyder and Daskin, 2006; Pintea et al.,2007; Hu and Raidl, 2008; Karapetyan and Gutin, 2011)

As far as we are aware, first integer linear programming formulation (ILPF) is proposed by Laporte and Nobert (1983). They consider the case where at least one node from each cluster must be visited. The authors concentrate on symmetric cost matrix and present an ILPF with an objective composed of travel and fixed costs. The symmetric GTSP has been studied in details by Fishetti, Gonzales and Toth (1995, 1997, 2002). Asymmetric version of the problem has been formulated by Laporte et al. (1987) and later by Noon and Bean (1991). When n is the number of the nodes in the graph, in all of these formulations, there are $O(n^2)$ binary variables but the number of the constrains grows exponentially with respect to n, which cause serious difficulties to solve such models directly by using any optimizer.

According to the best of our knowledge, first polynomial size ILPF for the GTSP is proposed by Kara et al. (2005). This formulation is a node based formulation and contains $O(n^2)$ decision variables and $O(n^2)$ constraints. Thus, we will not consider the older exponential size formulations. Later, Kara and Demir (2006) propose a new arc based polynomial size formulation with $O(n^2)$ constraints and decision variables. They computationally show that the node based formulation proposed by Kara et al. (2005) is superior to their arc based formulation in terms of linear programming relaxation and CPU time. Pop (2007), reviews the existing formulations for the GTSP and proposes new node and arc based formulations. Kara et al. (2009), adapt Pop's formulations to the symmetric and asymmetric cases, and computationally show that, the earlier node based formulation proposed by Kara et al. (2005) is superior to these formulations in terms of linear programming relaxations and CPU time. Consequently, the node based formulation proposed in 2005 is the outstanding formulation up to now. So, in this paper, for computational analysis we just consider the formulation proposed by Kara et al. (2005) as the comparable existing formulation.

There is a need to have formulations having polynomial number of integer variables and constraints, which is the main motivation and contribution of this paper. We propose two new polynomial size ILPFs for the GTSP. Then, we computationally analyze their performances by solving instances from the literature.

In section 2, we introduce a general ILPF for the GTSP, and clarify node and arc based formulations. In sections 3 and 4, the proposed formulations are presented. Preliminary computational results are summarized in section 5. The paper ends with a conclusion and further remarks in section 6.

2 General Formulation for the Generalized Traveling Salesman Problem

Let $G = (V, A)$ be a graph where $V = \{1, 2, \dots, n\}$ is the set of nodes and $A = \{(i, j)\colon i, j \in V, i \neq j\}$ is the set of *directed arcs* (or *edges*), and let c_{ij} be the travel distance (or cost or time) from node i to node j. Let $V_1, V_2, \dots, V_k$ be disjoint subsets of V such that union of these subsets equals to V. These subsets are called clusters. The GTSP is to find the tour that (*i*) starts from a node and visits exactly one node from each cluster and turns back to the starting node (*ii*) never visit a node more than once and (*iii*) has the minimum total tour length.

Associated with each arc, let x_{ij} be a binary variable equal to "1" if the traveler goes from node i to node j, and "0" otherwise. A general ILPF for the GTSP may be given as follows:

$$min \sum_{i \in V} \sum_{j \in V \setminus \{i\}} c_{ij} x_{ij} \qquad (1)$$

$subject\ to$

$$\sum_{i \in V_p} \sum_{j \in V \setminus V_p} x_{ij} = 1, \qquad p = 1, \dots, k \qquad (2)$$

$$\sum_{i \in V \setminus V_p} \sum_{j \in V_p} x_{ij} = 1, \qquad p = 1, \dots, k \qquad (3)$$

$$\sum_{j \in V \setminus \{i\}} x_{ji} - \sum_{j \in V \setminus \{i\}} x_{ij} = 0, \qquad \forall i \in V \qquad (4)$$

$$x_{ij} \in \{0,1\}, \qquad \forall (i,j) A \qquad (5)$$

$$Sub\ Tour\ Elimination\ Constraints \qquad (6)$$

Constraints (2) and (3) are degree constraints of each cluster, whereas, constraints (4) balance the inflow and outflow of each node. These constraints proposed by Noon and Bean (1991). Constraints (6) are sub tour elimination constraints, which must guarantee that the solution contains no illegal sub tours. Integrality constraints are given in (5).

Existing ILPFs of routing problems in the literature differ from each other with respect to the sub tour elimination constraints (SECs). In order to prevent illegal tours, i.e., eliminate sub tours, in most of the studies, additional decision variables are defined and then SECs of the formulation are developed. Such formulations may be divided into two groups according to these additional variables. We call a formulation **node based** if the additional variables are due to nodes, and **arc (flow)** based if they are due to arcs of the underlying graph.

For an ILPF of a routing problem, if the number of the constraints and/or decision variables of the formulation grows exponentially with respect to the number of the nodes, it's named as "*exponential size formulation*"; and if they grow polynomially, it's named as "*polynomial size formulation*".

The formulation of Kara et al. (2005) which is mentioned above and will be considered as the existing formulation in this study is given below. In this formulation u_p is the auxiliary variables defined as the visit number of cluster p and constraints (7) are the subtour elimination constraints of the formulation.

$$\min \sum_{i\in V} \sum_{j\in V\setminus\{i\}} c_{ij}x_{ij} \qquad (1)$$

subject to

(2), (3), (4), (5) and

$$u_p - u_q + k\sum_{i\in V_p}\sum_{j\in V_q} x_{ij} + (k-2)\sum_{i\in V_q}\sum_{j\in V_p} x_{ij} \le k-1, \quad p\neq q; p,q=2,\ldots,k \qquad (7)$$

$$u_p \ge 0, \qquad p=2,\ldots,k \qquad (8)$$

3 A New Node Based Formulation

In addition to the above auxiliary variables, let us define one more auxiliary variable as:

$$w_{pq} = \begin{cases} 1, & \text{if traveler goes from cluster p to cluster q} \\ 0, & \text{otherwise} \end{cases}$$

With these auxiliary variables, we propose a new node based formulation $F1$ for the GTSP as:

***F*1:**

$$\min \sum_{i\in V} \sum_{j\in V} c_{ij}x_{ij} \qquad (1)$$

subject to

(2), (3), (4), (5) and

$$w_{pq} = \sum_{i\in V_p}\sum_{j\in V_q} x_{ij}, \qquad p\neq q; p,q=1,\ldots,k \qquad (9)$$

$$u_p - u_q + (k-1)w_{pq} + (k-3)w_{qp} \le k-2, \qquad p\neq q; p,q=2,\ldots,k \qquad (10)$$

$$u_p - \sum_{\substack{q=2 \\ q\neq p}}^{k} w_{qp} \ge 1, \qquad p=2,\ldots,k \qquad (11)$$

$$u_p + (k-2)w_{1p} \le k-1, \qquad p=2,\ldots,k \qquad (12)$$

In $F1$, constraints (9) are expressions of the new variables by the previous decision variables. Constraints (10) are the subtour elimination constraints. Constraints (11) and (12) form the lower and upper bounds of the auxiliary variables. The validity of these constraints is shown in the propositions given below.

Proposition 1: The equalities given in (9) guarantee that w_{pq} equals to 1 if the traveler goes from cluster p to q, and 0 otherwise.

Proof: Equality given in (9) is the expression of w_{pq}'s as a function of the decision variables x_{ij}'s. In accordance with the constraints given in (2), (3), (4) and (5), right hand side of this equality is equal to 1 if traveler goes from cluster p to cluster q or equal to 0 otherwise. So, w_{pq}'s are binary derived decision variables corresponding to clusters. □

Proposition 2: Inequalities given in (10), (11) and (12) are valid for the GTSP.

Proof: With the new decision variables w_{pq}'s for clusters, constraints given in (11) and (12) initialize the auxiliary variable corresponding to the first cluster of the tour as equal to 1, i.e., u_p=1 when w_{1p}=1. As a result of the inequality given in (10), if w_{pq}=1 then u_q=u_p+1. So, u_p's of the clusters of the tour will cumulate and perform a step function. Thus, the auxiliary variables u_p's correspondence to the visit number of the p^{th} cluster and inequality (10) prohobits formation of illegal tours, i.e.,they are subtour elimination constraints of the formulation.□

Proposed formulation $F1$ has $|A|$ binary and (k^2-k) continuous variables, and $(n + 2k^2)$ constraints. So, $F1$ contains $O(n^2)$ binary variables and $O(n^2)$

ISBN: 978-1-61804-076-3

constraints., i.e., $F1$ is a polynomial size formulation for the GTSP.

We now propose another new polynomial formulation for the GTSP below.

4 A New Arc Based Formulation

With w_{pq} variables which are defined in $F1$, let us define a new auxiliary variable for the number of arcs on the tour as:

y_{pq}: the number of arcs which are occurred on the tour until the cluster q if traveler goes from cluster p to cluster q.

With these new variables, a new arc based formulation, $F2$, is proposed for the GTSP as:

***F*2:**

$$\min \sum_{i\in V}\sum_{j\in V} c_{ij}x_{ij} \qquad (1)$$

subject to

(2), (3), (4), (5), (9) and

$$y_{pq} \leq k^* w_{pq}\ , \qquad p\neq q; p=2,\ldots,k; q=1,\ldots,k \qquad (13)$$

$$y_{1p} = w_{1p}\ , \qquad p=2,\ldots,k \qquad (14)$$

$$\sum_{q\neq p} y_{pq} - \sum_{r\neq p} y_{rp} = 1\ , \quad p=2,\ldots,k \qquad (15)$$

In this formulation, constraints (13) and (14) give the lower and upper bounds of the connections between the clusters. Constraints (15) prevent the sub tours. The validity of these constraints is given in proposition 3 below.

Proposition 3: Inequalities given in (13), (14) and (15) are valid for the GTSP.

Proof: Inequality given in (13) gives upper bound for the y_{pq}'s and guarantee that if w_{pq}=0 then y_{pq}=0. Equation (14) implies that, if the traveler goes from the first cluster to the cluster p then w_{1p}=1 and we will initialize y_{1p}=1. If w_{pq}=1, equation given in (15) implies that number of the arcs on the tour up to cluster p is increased by one for the number of the arcs up to cluster q. So, equation (15) will perform a step function between the y_{pq}'s of the adjacent clusters. Thus, it will prohibit formation of illegal tours, i.e., they are sub tour elimination constraints of the formulation.□

Proposed formulation $F2$ has |A| binary and ($2k^2$-$2k$) continuous variables, and (n+$2k^2$+k-1) constraints. So, $F2$ contains $O(n^2)$ binary variables and $O(n^2)$ constraints., i.e., $F2$ is a polynomial size formulation for the GTSP.

5 Preliminary Computational Analyses

The linear programming (LP) formulation obtained by omitting all integrality or 0-1 constraints on variables is called the *LP relaxation* of the formulation in the decision models. The computational analyses are made in terms of the solution times and LP relaxation values (LPR values) of the models. 19 asymmetric and 32 symmetric GTSP instances are taken from TSPLIB (http://www.cs.rhul.ac.uk/home/zvero/GTSPLIB). These problems are solved with the CPLEX 12.0 solver by a computer having Intel Core i5 processor with 4 cores, 4 GB of RAM and 276 GB of SSD drive. The solution times are limited to 7200 seconds. The benchmarking problems are solved with the existing (Kara et al., 2005) and proposed formulations $F1$ and $F2$. Preliminary results are summarized in the following part.

Asymmetric Problems: There does not exist significant differences between the linear programming relaxations. All three formulations produced optimal solutions of all 19 test problems. Mean of the solution times of the existing formulation is 84.26 seconds while it is 29.34 seconds for $F1$ and 96.53 seconds for $F2$. So, mean of the solution time of $F1$ is superior to the existing and new arc based formulation. Consequently, for the asymmetric GTSP, if one wants to use a formulation, we certainly propose formulation $F1$.

Symmetric Problems: Within predetermined time limit 7200 seconds, existing formulation solved 15 problems optimally while $F2$ solved 21 and $F1$ reached optimal solution of the 22 test problems. So, both new formulations are superior to the existing one. Mean of the fluctuation of the linear programming relaxations from the optimal values of 22 problems is found as 0,345 for $F1$ and 0,37 for $F2$. So, node based formulation produces a little bit better lower bound than arc based formulation of the GTSP. Mean solution time of 21 problems that both $F1$ and $F2$ obtained optimal solutions shows that $F1$ is better than $F2$.

ISBN: 978-1-61804-076-3

6 Conclusions and Further Remarks

Two polynomial size integer programming formulations for the generalized traveling salesman problem are presented. The first formulation, designated as $F1$, is node based and the second one, designated as $F2$ is arc based. Both formulations have $O(n^2)$ binary variables and $O(n^2)$ constraints. Preliminary computational analysis show that, for both asymmetric and symmetric instances, performance of $F1$ is better than the existing node based formulation and $F2$ in terms of CPU time. There does not exist significant differences between the linear programming relaxations of the formulations. Thus we recommend to use of our node based formulation for the GTSP.

To develop valid inequalities for $F1$ and $F2$ and conduct detailed computational analysis on the proposed formulations, and to adapt the formulations to multiple traveler case seem further research areas.

References:

[1] Aziz I.A., Haron N., Mehat M., Jung L.T., Mustapa A.N., Akhir E.A.P., Solving Traveling Salesman Problem on Cluster Compute Nodes, *WSEAS TRANSACTIONS on COMPUTERS,* Issue 6, Volume 8, June 2009, 1020-1029.

[2] Ben-Arieh D., Gutin G., Penn M., Yeo A., Zverovitch A., Transformations of Generalized ATSP into ATSP, *Operations Research Letters*, 31 (5), 2003, 357-365.

[3] Dimitrijevic V., Saric Z., An Efficient Transformation of the Generalized Traveling Salesman Problem into Traveling Salesman Problem on Digraphs, *Information Sciences,* 102**,** 1997, 105–110.

[4] Dinh T.H., Mamun A.A., A Combination of Clustering Algorithms with Ant Colony Optimization for Large Clustered Euclidean Travelling Salesman Problem, *WSEAS Transactions on Systems, 3, no.3*, 2004, 1221-1227.

[5] Fischetti M., Gonzales J.J.S., Toth P., The Symmetric Generalized Traveling Salesman Polytope, *Networks* 26, 1995, 113–123.

[6] Fischetti M., Gonzales J.J.S., Toth P., A Branch-and-cut Algorithm for the Symmetric Generalized Traveling Salesman Problem, *Oper. Res.* 45, 1997, 378–394.

[7] Fischetti M., Gonzales J.J.S., Toth P., The Generalized Traveling Salesman and Orienteering Problem, In: G. Gutin and A.P. Punnen, Editors, *Traveling Salesman Problem and its Variations*, Kluwer Academic Publishers, Dordrecht, 2002.

[8] Gutin G., Punnen A. P. (Edit), *The Traveling Salesman Problem and its Variations*, Kluwer Academic Publishers, London, 2002.

[9] Hu B., Raidl G. R., Effective Neighborhood Structures for the Generalized Traveling Salesman Problem, J.van Hemert and C. Cotta (Eds.): *EvoCOP 2008, LNCS* 4972, 2008, 36-47.

[10] Huang H., Yang X., Hao,Z., Wu, C.,Liang,Y., Zhao,X., Hybrid Chromosome Genetic Algorithm for Generalized Traveling Salesman Problems, *LNCS* 3612, 137-140.

[11] Kara İ**.**, Derya T., Demir E., Bektaş T., An Integer Programming Formulation for Generalized Traveling Salesman Problem (In Turkish), YA/EM 2005, İstanbul.

[12] Kara İ., Demir E., Integer Programming Formulations for Generalized Traveling Salesman Problem (In Turkish), YA/EM 2006, İzmit.

[13] Kara İ., Güden H., Öztürk A., Seyhan A., Computational Results of the Polynomial Size Formulations of the Generalized Traveling Salesman Problem (In Turkish), 10th Econometric and Statistics Symposium, Erzurum, 2009.

[14] Karapetyan D., Gutin G., Lin-Kernighan Heuristic Adaptations for the Generalized Traveling Salesman Problem, *European Journal of Operational Research* 208, 2011, 221-232.

[15] Laporte G., Palekar U., Some Applications of the Clustered Travelling Salesman Problem, *Journal of the Operational Research Society*, 53, 2002, 972-976.

[16] Laporte G., Asef-Vaziri A., Sriskandarajah C., Some Applications of the Generalized Traveling Salesman Problem, *Journal of the Operational Research Society*, 47, 1996, 1461-1467.

[17] Laporte G., Mercure H., Nobert Y., Generalized Travelling Salesman Problem Through *n* sets of Nodes: the Asymmetrical Cases, *Discrete Appl. Math.* 18, 1987, 185–197.

[18] Laporte G., Nobert Y., Generalized Traveling Salesman Problem Through n sets of Nodes: an Integer Programming Approach, INFOR 21, 1983, 61-75.

[19] Laporte G., Semet F., Computational Evaluation of a Transformation Procedure for the Symmetric Generalized Traveling Salesman Problem, *INFOR* 37, 1999, 114–120.

[20] Lawler E.L., Lenstra J.K., Rinnooy Kaan A., Shmoys D.B. (Edit.), *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization,* Wiley, NewYork, 1985.
[21] Lien Y., Ma E., Wah B.W.S., Transformation of the generalized traveling salesman problem into the standard traveling salesman problem, *Information Sci.* 74, 1993, 177–189.
[22] Noon C. E., Bean J. C., A Lagrangian based approach for the asymmetric generalized traveling salesman problem, *Operations Research*, 39, 1991, 623-632.
[23] Noon C. E., Bean J. C., An Efficient Transformation of the Generalized Traveling Salesman Problem, INFOR 31(1), 1993, 39-44.
[24] Pintea C., Pop P.C., Dumitrescu D., An Ant-based Technique for the Dynamic Generalized Traveling Salesman Problem, *Proceedings of the 7th WSEAS International Conference on Systems Theory and Scientific Computation*, Athens, Greece, August 24-26, 2007, 255-259.
[25] Pop P. C., New Integer Programming Formulations of the Generalized Traveling Salesman Problem, *American Journal of Applied Sciences*, 4 (11), 2007, 932-937.
[26] Pop P.C., Sitar C.P., Pop N., Zelina I., A New Approach to Generalized Network Design Problems, *Proceedings of the 6th WSEAS Int. Conf. on Electronics, Hardware, Wireless and Optical Communications,* Corfu Island, Greece, 2007.
[27] Renaud J., Boctor F., F., An Efficient Composite Heuristic for the Symmetric Generalized Traveling Salesman Problem, *European Journal of Operational Research* 108 (3), 1998, 571-584.
[28] Snyder L.V., Daskin M.S., A Random-key Genetic Algorithm for the Generalized Traveling Salesman Problem, *European Journal of Operational Research*, 174, 2006, 38-53.

ISBN: 978-1-61804-076-3