A Bilevel Model and Solution Algorithm for a Freight Tariff-Setting Problem

LUCE BROTCORNE

LAMIH, Université de Valenciennes, Le Mont Houy BP311, 59304 Valenciennes France, ISRO and SMG, Université libre de Bruxelles CP 210/01, Boulevard du Triomphe B-1050 Bruxelles, Belgium, and CRT, Université de Montréal, C.P. 6128 Succursale Centre-Ville, Montréal H3C 3J7, Canada

MARTINE LABBÉ

ISRO and SMG, Université libre de Bruxelles CP 210/01, Boulevard du Triomphe B-1050 Bruxelles, Belgium

PATRICE MARCOTTE

CRT and DIRO, Université de Montréal, C.P. 6128 Succursale Centre-Ville, Montréal H3C 3J7, Canada

GILLES SAVARD

GERAD and DMGI, École Polytechnique de Montréal, CP 6079 Succursale Centre-Ville, Montréal H3C 3A7, Canada

We consider a bilevel programming formulation of a freight tariff-setting problem where the leader consists in one among a group of competing carriers and the follower is a shipper. At the upper level, the leader's revenue corresponds to the total tariffs levied, whereas the shipper minimizes its transportation cost, given the tariff schedule set by the leader. We propose for this problem a class of heuristic procedures whose relative efficiencies, on small problem instances, could be validated with respect to optimal solutions obtained from a mixed integer reformulation of the mathematical model. We also present numerical results on large instances that could not be solved to optimality by an exact method.

This work is devoted to a tariff-setting problem involving two decision makers acting non-cooperatively and in a sequential way. We focus our attention on a freight transportation application. In this context, a shipper company (the follower) is set to ship a prescribed amount of goods from origin nodes to its customers at minimum cost. Supply at the origin nodes and demand from the customers are both assumed to be known and fixed. Hence, for a given tariff schedule, the shipper's problem consists of satisfying demand at the lowest possible cost. The ensuing flow repartition is obtained by solving a standard transshipment problem where the tariffs are added to the initial arc costs. This is the lower level problem.

At the upper level, a given carrier (the leader) strives to maximize its revenues by setting optimal tariffs on the subset of arcs in its control. This carrier assumes no reaction from its competitors, but explicitly takes into account the reaction of the shipper company to its price schedule. The remaining carriers may represent different transportation modes or agents within a mode. We explicitly divide the freight rates on the links controlled by the leader carrier into two parts: the carrier's operating costs and an additional tariff. The unit profit associated with a given link is obtained by subtracting the unit operating cost from the unit freight rate.

This sequential and non-cooperative decision-

Transportation Science, © 2000 INFORMS
Vol. 34, No. 3, August 2000

0041-1655/00/3403-0289 $05.00
pp. 289–302, 1526-5447 electronic ISSN

making process can be adequately represented as a bilevel program where the upper level objective is bilinear and the lower level's bilinear objective is actually a linear program parameterized by the upper level decision vector. This constitutes a particular case of the general taxation model introduced by LABBÉ, MARCOTTE, and SAVARD (1998). This paper focuses on algorithms aimed at solving large-scale instances of this NP-hard problem.

Until recently, the literature on freight networks mainly focused on the freight network equilibrium problem, without addressing the problem of what tariffs carriers should charge (cf., for example, HARKER, 1988; MARCOTTE, 1987). However, with the current emphasis on deregulation, it makes sense to consider the carriers as active players in a game. Game theory provides a framework for analyzing the interactions between the carriers. More precisely, the routing of freight flows between supply and demand sites, and the tariffs and service levels set by the carriers, are determined by assuming that the shippers minimize their respective transportation costs and that the carriers maximize their respective profits, taking into account the response of the shippers. For instance, FRIESZ, GOTTFRIED, and MORLOK (1986) propose a sequential shipper–carrier model. Shippers first select sites to purchase goods and the transportation agents who will ship the goods to the desired destinations to minimize their cost. This determines the transportation demand. Then, carriers respond by routing freight on the links of the network to minimize the total operation cost. In this model, tariffs are a constant fraction of the cost incurred by carriers and consequently not a decision variable for the agents. FISK (1986) proposes a model where one or several competing carriers achieve, through the selection of tariffs and levels of service, a Nash equilibrium of the resulting oligopolistic market. The demand side corresponds to production plants selecting carriers and routes to move their products to destinations at lowest cost. No procedure has been proposed to solve this difficult problem, which subsumes ours.

More recently, HURLEY AND PETERSEN (1994a,b), have described two models where carriers first determine tariffs and, subsequently, shippers select the production levels and a coalition of carriers that will transport the production from origins to destinations at minimal cost. Although their non-cooperative models have a bilevel flavor, they both reduce to the problem of maximizing the joint profit of shippers and carriers. This single-agent problem possesses the structure of a traffic assignment problem, for which several efficient algorithms are known. Finally, the distribution of the joint profit among the agents of the system is obtained by solving a linear program. A key feature of their models is the functional form of the tariffs. Rather than assuming that tariffs are proportional to the volume of shipments (linear tariffs), which may fail to be optimal and is inconsistent with usual practice (cf. WILSON, 1993; HURLEY and PETERSEN, 1994b; and TIROLE, 1989), they analyze nonlinear, two-part tariffs. These involve a fixed cost for an initial shipment, and, next, a smaller constant price for additional shipments. The authors present numerical results on small, simplified networks.

Our approach is different in that we focus on the tariff side rather than on the allocation of freight flows among the carriers. Also, in contrast to Fisk (1986), Friesz, Gottfried, and Morlok (1986), Harker (1988), and Hurley and Petersen (1994a,b), who consider multicommodity flows, we only address the single-commodity transshipment problem. This is the case of lower level firms whose demand for transportation depends solely on the location of its supply and demand sites. Such a situation occurs in the distribution of coal supplies in the gas industry, the distribution of containers or seasonal products, etc. Finally, note that the model described in this paper is not bound to a specific, predefined analytical form of the tariff function. The tariff is determined with respect to the actual shipments. Should these shipments vary, then the profit-maximizing tariffs should be reevaluated.

This model could also be part of the yield management process in the rail or airline industry, where the pricing strategies must take into account the competitors' supply and fare structures. An underlying assumption of our approach is that the leader is not a dominant player of the market. This implies that it is reasonable to assume that total demand is not influenced by leader's prices (this could be the case of the trucking industry) and that the competition does not react in the short term to the leader's prices. In that sense, our model is not oligopolistic.

The remainder of this paper is organized as follows. In Section 1, we give a mathematical formulation of the freight transportation problem. Four primal–dual heuristic procedures are described in Section 2. In Section 3 we propose two strategies aimed at improving upon solutions initially obtained by the heuristics. Computational results are presented in Section 4, followed by a conclusion in Section 5.

1. FORMULATION OF THE FREIGHT TARIFF-SETTING PROBLEM

WE CONSIDER A transportation network based on the underlying graph $G = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ (of cardinality n) denotes the set of nodes, and $\mathcal{A}$ (of cardinality m) the set of arcs. A node represents either a supply site, a demand site, or the end of an arc on which goods are carried. The arc set $\mathcal{A}$ of the network G is partitioned into two subsets $\mathcal{A}_1$ and $\mathcal{A}_2$, where $\mathcal{A}_1$ denotes the set of links operated by the leader carrier, and $\mathcal{A}_2$ the set of links operated by its competitors. A service between two nodes operated by both the leader and its competitors would be represented in the network by two parallel links, respectively members of the sets $\mathcal{A}_1$ and $\mathcal{A}_2$.

With each tariff arc a of $\mathcal{A}_1$, we associate a freight rate composed of a fixed cost c_a representing the unit traversal cost of the arc, and an additional tariff T_a to be determined by the leader carrier. A free arc a of $\mathcal{A}_2$ only bears a unit cost d_a, which is outside the control of the leader, and might include both a fixed cost and the cost charged by other carriers as well. We denote by $b \in R^n$ the (fixed) demand for transportation, with the tacit assumption that supply corresponds to negative demand. Under this convention, demand at transshipment nodes, i.e., nodes that are neither origins nor destinations, is zero. Furthermore, and without loss of generality, we assume that each component of the demand vector b is integer valued.

For given freight rates T_a set by the leader, the shipper's distribution problem is a transshipment problem. Its optimal (basic) solutions will consist of the unique assignment of the demand flow on some subtree of the graph G. We make the assumption that, among trees of equal costs for the follower, goods will be shipped on a tree that maximizes the leader's profit. This is a standard assumption underlying the bilevel model that can be substantiated by noting that a nearly optimal solution can be achieved through an arbitrarily small tariff reduction.

TABLE I
Notation

$\mathcal{N}$	Node set
n	Number of nodes
$\mathcal{A}$	Arc set
$\mathcal{A}_1$	Set of tariff arcs
$\mathcal{A}_2$	Set of free arcs
m	Number of arcs
m_1	Number of tariff arcs
m_2	Number of free arcs
i^-	$\{(j, i) \in \mathcal{A}: j \in \mathcal{N}\}$
i^+	$\{(i, j) \in \mathcal{A}: j \in \mathcal{N}\}$
b_i	Demand at node i
c_a	Cost per unit of $a \in \mathcal{A}_1$
d_a	Cost per unit of $a \in \mathcal{A}_2$
T_a	Tariff per unit on $a \in \mathcal{A}_1$
x_a	Flow on arc $a \in \mathcal{A}_1$
y_a	Flow on arc $a \in \mathcal{A}_2$

Based on this notation (see also Table I), the freight tariff-setting problem (FTSP) can be expressed as a bilevel program with bilinear objectives and linear constraints, namely,

FTSP

$$\max_{T,x} \sum_{a\in\mathcal{A}_1} T_a x_a$$

$$\min_{x,y} \sum_{a\in\mathcal{A}_1} (c_a + T_a) x_a + \sum_{a\in\mathcal{A}_2} d_a y_a$$

subject to

$$\sum_{a\in i^-\cap\mathcal{A}_1} x_a + \sum_{a\in i^-\cap\mathcal{A}_2} y_a - \sum_{a\in i^+\cap\mathcal{A}_1} x_a - \sum_{a\in i^+\cap\mathcal{A}_2} y_a = b_i \quad \forall i \in \mathcal{N},$$

$$x_a \geqslant 0 \quad \forall a \in \mathcal{A}_1, \qquad y_a \geqslant 0 \quad \forall a \in \mathcal{A}_2, \tag{1}$$

where i^+ (respectively i^-) denotes the set of arcs having node i as its tail (respectively head). The leader's total profit is obtained by summing, over all tariff arcs, the product of the unit profit with the corresponding arc flow, thus naturally yielding a bilinear objective. At the lower level, the shipper firm satisfies the demand at lowest cost, by solving a linear transshipment problem whose objective is parameterized by the leader's tariff, thus yielding a bilinear lower level objective for the bilevel program.

This fits the framework of the general taxation problem studied by Labbé, Marcotte and Savard (1998). If one incorporates into the leader's objective the reaction function of the follower, the resulting single level objective is not a convex, indeed not even a continuous, function of the tariff vector T. However, because the derived profit mapping is upper semicontinuous, one can conclude that the set of optimal solutions to the freight tariff-setting problem is nonempty if the profit is bounded from above.

Throughout the paper, we assume that there exists a feasible transportation schedule that uses none of the arcs in control of the leader. This assumption is both necessary and sufficient for the leader's profit to be bounded from above. This upper bound will be finite whenever the fixed part of the arc costs is non-negative, thus preventing the occurrence of profitable negative cost circuits in an optimal leader's solution. More precisely, let us denote

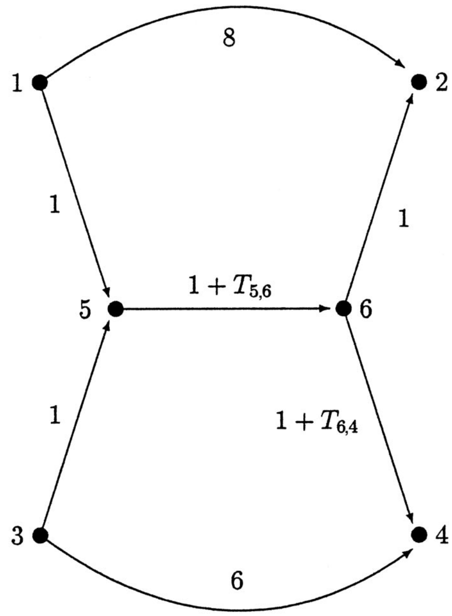

Fig. 1. Example with negative tariffs.

by $\tau(T)$ the lower level transshipment cost induced by a tariff schedule T. It is clear that the follower will never accept a cost higher than $\tau(\infty)$, corresponding to a solution with no flow on the tariff arcs. Now, consider an optimal solution vector (x^*, y^*, T^*). Because $T^*x^* \geqslant 0$, we have

$$(c + T^*)x^* + dy^* \geqslant cx^* + dy^* \geqslant \min_{x,y} cx + dy = \tau(0),$$

i.e., that the follower's cost at an optimal solution will be at least $\tau(0)$, the cost of a lower level solution corresponding to null tariffs. Hence, an upper bound on the leader's profit is given by the difference $\tau(\infty) - \tau(0)$. However, as noticed by Labbé, Marcotte and Savard (1998), this bound is not always reached, even in the single-origin–single-destination case.

We do not impose sign constraints on tariffs, because negative tariffs can induce compensating effects that result in higher profits for the leader. For instance, consider the example of Figure 1, where arcs (5, 6) and (6, 4) are subject to tariffs, and the demand for transportation is given by

$$b = (-1, 1, -1, 1, 0, 0).$$

In this particular case, compensating interactions are present and the optimal solution with profit 8 is reached with tariff values $T_{5,6} = 5$ and $T_{6,4} = -2$.

The above bilevel program can be formulated as a mixed integer linear program. This formulation will be used to solve exactly small instances of the taxation problem FTSP and thus to evaluate the quality of the solutions produced by the proposed heuristics. This formulation is obtained by first substituting, for the lower level linear program, its necessary and sufficient optimality conditions, thus yielding the single level equivalent program.

BILIN

$$\max_{T,x,y,\lambda} \sum_{a\in\mathcal{A}_1} T_a x_a$$

subject to

$$\sum_{a\in i^-\cap\mathcal{A}_1} x_a + \sum_{a\in i^-\cap\mathcal{A}_2} y_a - \sum_{a\in i^+\cap\mathcal{A}_1} x_a - \sum_{a\in i^+\cap\mathcal{A}_2} y_a = b_i \quad \forall i \in \mathcal{N}$$

$$x_a \geqslant 0 \quad \forall a \in \mathcal{A}_1$$
$$y_a \geqslant 0 \quad \forall a \in \mathcal{A}_2$$
$$\lambda_j - \lambda_i \leqslant c_{i,j} + T_{i,j} \quad \forall (i, j) \in \mathcal{A}_1$$
$$\lambda_j - \lambda_i \leqslant d_{i,j} \quad \forall (i, j) \in \mathcal{A}_2$$
$$\sum_{a\in\mathcal{A}_1} (c_a + T_a)x_a + \sum_{a\in\mathcal{A}_2} d_a y_a = \sum_{i\in\mathcal{N}} \lambda_i b_i.$$

If each arc flow were associated with a single origin–destination pair, the objective and the nonlinear constraint (sixth constraint) of the above nonlinear program could quite easily be linearized. Unfortunately, this is not the case. However one can achieve the same goal through the binary expansion of the flow variables. More precisely, because

$$0 \leqslant x_a \leqslant \sum_{i\in\mathcal{N}:b_i>0} b_i,$$

one can write

$$x_a = \sum_{k=0}^{\bar{k}} 2^k z_a^k,$$

where $z_a^k \in \{0, 1\}$ for all k such that $0 \leqslant k \leqslant \bar{k}$, with $\bar{k} = \lfloor \log_2(\Sigma_{i\in\mathcal{N}:b_i>0}(b_i + 1)) \rfloor$. Next, one introduces a tariff variable $T_a^k = T_a z_a^k$ for all k such that $k \in \mathcal{K}$ where $\mathcal{K} = [1, \ldots, \bar{k}]$, and we incorporate into the model the constraints

$$-Mz_a^k \leqslant T_a^k \leqslant Mz_a^k \quad \forall a \in \mathcal{A}_1 \quad \forall k \in \mathcal{K}$$
$$-M(1 - z_a^k) \leqslant T_a^k - T_a \leqslant M(1 - z_a^k) \quad \forall a \in \mathcal{A}_1 \quad \forall k \in \mathcal{K}$$
$$z_a^k \in \{0, 1\} \quad \forall a \in \mathcal{A}_1 \quad \forall k \in \mathcal{K},$$

where M is some constant arbitrary large with respect to data values. These modifications yield the

following mixed integer programming formulation of the freight tariff-setting problem:
MIP

$$\max_{T,y,z,\lambda} \sum_{a\in\mathcal{A}_1} \sum_{k=0}^{\bar{k}} 2^k T_a^k$$

subject to

$$\sum_{a\in i^-\cap\mathcal{A}_1} \sum_{k=0}^{\bar{k}} 2^k z_a^k + \sum_{a\in i^-\cap\mathcal{A}_2} y_a - \sum_{a\in i^+\cap\mathcal{A}_1} \sum_{k=0}^{\bar{k}} 2^k z_a^k - \sum_{a\in i^+\cap\mathcal{A}_2} y_a = b_i \quad \forall i \in \mathcal{N}$$

$$\lambda_j - \lambda_i \leqslant c_{i,j} + T_{i,j} \quad \forall (i, j) \in \mathcal{A}_1$$

$$\lambda_j - \lambda_i \leqslant d_{i,j} \quad \forall (i, j) \in \mathcal{A}_2$$

$$\sum_{a\in\mathcal{A}_1} \left(c_a \sum_{k=0}^{\bar{k}} 2^k z_a^k + \sum_{k=0}^{\bar{k}} 2^k T_a^k \right) + \sum_{a\in\mathcal{A}_2} d_a y_a = \sum_{i\in\mathcal{N}} \lambda_i b_i$$

$$-M z_a^k \leqslant T_a^k \leqslant M z_a^k \quad \forall a \in \mathcal{A}_1 \quad \forall k \in \mathcal{K}$$

$$-M(1 - z_a^k) \leqslant T_a^k - T_a \leqslant M(1 - z_a^k) \quad \forall a \in \mathcal{A}_1 \quad \forall k \in \mathcal{K}$$

$$z_a^k \in \{0, 1\} \quad \forall a \in \mathcal{A}_1 \quad \forall k \in \mathcal{K}$$

$$y_a \geqslant 0 \quad \forall a \in \mathcal{A}_2.$$

2. EFFICIENT HEURISTIC PROCEDURES

THE FREIGHT TARIFF-SETTING problem can be formulated as a single level bilinear program with disjoint constraints. Due to the size of these problems (thousands of variables and constraints), a direct bilinear programming approach (AUDET et al., 1999 or THIEU, 1988) is not suitable. This justifies the development of metaheuristic procedures that explicitly take into account the network structure of the problem.

In this section, we describe four primal–dual heuristic procedures. A first group of algorithms is inspired by a primal–dual heuristic proposed by GENDREAU, MARCOTTE, and SAVARD (1996) for solving linear bilevel programs. This scheme can be applied in two different ways to each formulation, which makes for a total of four algorithms. Symmetrically, a Gauss–Seidel-based algorithm can be implemented in two different ways for each formulation. This makes for a grand total of eight conceptual heuristic procedures. Four of these have been implemented and computationally tested in the present paper.

The algorithms address a reformulation of FTSP as a single-level bilinear. For convenience, let us express FTSP in vector-matrix notation:

FTSP
$$\max_{T,x} Tx$$
$$\min_{x,y} (c + T)x + dy$$
$$\text{subject to} \quad A_1x + A_2y = b$$
$$x, y \geqslant 0,$$

where $A_1 \in R^{n\times m_1}$ denotes the node–arc incidence matrix of the subnetwork composed of tariff arcs and $A_2 \in R^{n\times m_2}$ denotes the incidence matrix corresponding to the free part of the network. Using this notation, the single level equivalent of the FTSP takes the form,

$$\begin{aligned} &\max_{T,x,y,\lambda} Tx \\ &\text{subject to} \quad A_1x + A_2y = b \\ &\qquad x, y \geqslant 0 \\ &\qquad \lambda A_1 \leqslant c + T \\ &\qquad \lambda A_2 \leqslant d \\ &\qquad (c + T)x + dy - \lambda b = 0. \end{aligned} \tag{3}$$

Next, we penalize the last constraint of Eq. 3 stating the equality of the primal and dual objectives. The left-hand side of this constraint is nonnegative whenever (x, y) and λ are feasible for the primal and dual problems, respectively. This yields the bilinear program,

$$\begin{aligned} &\max_{T,x,y,\lambda} Tx - M_1((c + T)x + dy - \lambda b) \\ &\text{subject to} \quad A_1x + A_2y = b \\ &\qquad x, y \geqslant 0 \\ &\qquad \lambda A_1 \leqslant c + T \\ &\qquad \lambda A_2 \leqslant d, \end{aligned} \tag{4}$$

where $M_1 > 0$. This penalty scheme is exact in the sense that there exists a finite value M^* such that any optimal solution of Eq. 3 is also optimal for BILIN, and vice versa, whenever $M_1 \geqslant M^*$ (see Labbé, Marcotte and Savard (1998) or, in the context of linear bilevel programs, ANANDALINGAM and WHITE (1990)).

For given vectors x, y, and λ, the above problem reduces to

$$\max_{T} (1 - M_1)Tx$$
$$\text{subject to} \quad \lambda A_1 - c \leqslant T. \qquad (5)$$

Because the flow vector x is nonnegative, an optimal solution of this linear program is obtained by setting T to $+\infty$ if $1 - M_1 > 0$ (yielding an infinite profit), to its lower bound $T = \lambda A_1 - c$ if $1 - M_1 < 0$, and to an arbitrary feasible vector if $1 - M_1 = 0$. The only interesting situation occurs when M_1 is strictly larger than 1, in which case problem 4 can be rewritten as the bilinear program,

PEN1

$$\max_{x,y,\lambda}(\lambda A_1 - c)x - M_1(\lambda A_1 x + dy - \lambda b) \qquad (6)$$
$$\begin{aligned} \text{subject to} \quad & A_1x + A_2y = b \\ & x, y \geqslant 0 \\ & \lambda A_2 \leqslant d. \end{aligned}$$

A slightly different reformulation of problem FTSP as a single level bilinear program, proposed by Labbé, Marcotte, and Savard (1998), may be obtained by expressing the optimality conditions of the lower level program through complementarity slackness. This yields the mathematical program

$$\max_{T,x,y,\lambda} Tx$$
$$\begin{aligned} \text{subject to} \quad & A_1x + A_2y = b \\ & x, y \geqslant 0 \\ & \lambda A_1 \leqslant c + T \\ & \lambda A_2 \leqslant d \\ & (d - \lambda A_2)y = 0 \\ & (c + T - \lambda A_1)x = 0. \end{aligned} \qquad (7)$$

Then, from complementarity slackness and primal feasibility, the objective function of problem 7 can be written as

$$\begin{aligned} Tx &= (\lambda A_1 - c)x \\ &= \lambda(b - A_2y) - cx \\ &= \lambda b - cx - dy. \end{aligned}$$

Thus, problem 7 is equivalent to

$$\max_{T,x,y,\lambda} \lambda b - (cx + dy)$$
$$\begin{aligned} \text{subject to} \quad & A_1x + A_2y = b \\ & x, y \geqslant 0 \\ & \lambda A_1 \leqslant c + T \\ & \lambda A_2 \leqslant d \\ & (d - \lambda A_2)y = 0 \\ & (c + T - \lambda A_1)x = 0. \end{aligned} \qquad (8)$$

Note the role played by the toll vector T in the above formulation: T is not part of the objective and, for any values taken by the other variables, there always exists a value of T that satisfies the third and sixth constraints of Eq. 8. Actually, after the removal of those two constraints and the optimization of the resulting problem over the variables x, y, and λ, an optimal solution to Eq. 8 can be recovered by simply setting

$$T = \lambda A_1 - c.$$

Actually, the equality only needs to hold for the components of the flow vector x that are strictly positive. For those components that are equal to zero, one is free to select any value of T that is sufficiently large, i.e.,

$$T_{i,j} \geqslant (\lambda A_1 - c)_{i,j},$$

if $x_{i,j} = 0$. In practice, it might be convenient to set $T_{i,j}$ to $+\infty$ whenever $x_{i,j} = 0$, to forbid the usage of these arcs. This yields the mathematical program,

$$\max_{T,x,y,\lambda} \lambda b - (cx + dy)$$
$$\begin{aligned} \text{subject to} \quad & A_1x + A_2y = b \\ & x, y \geqslant 0 \\ & \lambda A_2 \leqslant d \\ & (d - \lambda A_2)y = 0. \end{aligned} \qquad (9)$$

Finally, we penalize the complementarity constraint of 9 into the objective to obtain the bilinear problem,

PEN2

$$\max_{x,y,\lambda} \lambda b - (cx + dy) - M_2(d - \lambda A_2)y \qquad (10)$$
$$\begin{aligned} \text{subject to} \quad & A_1x + A_2y = b \\ & x, y \geqslant 0 \\ & \lambda A_2 \leqslant d, \end{aligned}$$

where M_2 is some positive constant. As was the case for the formulation PEN1, PEN2 is equivalent to FTSP whenever the penalty parameter M_2 is sufficiently large.

All the proposed heuristic procedures are based on two simple underlying principles:

- Approximate the original bilevel program by a sequence of penalized problems with increasing levels of penalty parameters;
- Solve (approximately) each penalized problem by iteratively solving a sequence of linear programs.

The first principle allows to start from a best solution for the leader, from which to move toward feasible solutions, i.e., solutions that satisfy the follower's optimality conditions. This best solution, which provides an upper bound on the leader's profit, corresponds to setting the penalty parameter to zero (actually to one in the case of heuristic 1), yielding the linear program,

$$\begin{aligned} \max_{x,y,\lambda} \ & \lambda b - (cx + dy) \\ \text{subject to} \ & A_1 x + A_2 y = b \\ & x, y \geqslant 0 \\ & \lambda A_2 \leqslant d. \end{aligned}$$

This linear program decomposes into two subprograms, one with respect to the (x, y) vector, and the other with respect to the dual vector λ. One can thus write this upper bound as

$$\begin{array}{lcl} \max_{\lambda} \lambda b & - & \min_{x,y} cx + dy \\ \text{subject to } \lambda A_2 \leqslant d & & \text{subject to } A_1 x + A_2 y = b \\ & & \qquad x, y \geqslant 0, \end{array}$$

or, replacing the left linear program by its equivalent dual,

$$\begin{array}{lcl} \min_{y'} dy' & & \min_{x,y} cx + dy \\ \text{subject to } A_2 y' = b & - & \text{subject to } A_1 x + A_2 y = b \\ \qquad y' \geqslant 0 & & \qquad x, y \geqslant 0. \end{array}$$

One observes that the LP on the left has to an optimal lower level solution that uses none of the toll arcs, whereas the solution of the right LP is a lower level solution corresponding to a zero toll vector T. This is exactly the upper bound alluded to in the previous section.

The second principle allows replacement of a difficult problem by a sequence of easy problems. Note that the constraint sets of the bilinear programs PEN1 and PEN2 are separable in the dual vector λ on the one hand, and in the primal vector (x, y) on the other hand. This suggests iteratively solving for the primal vector (x, y) and the dual vector λ. Note that one could also have iterated based on the partition $\{(\lambda, y), x\}$, although this strategy was not retained.

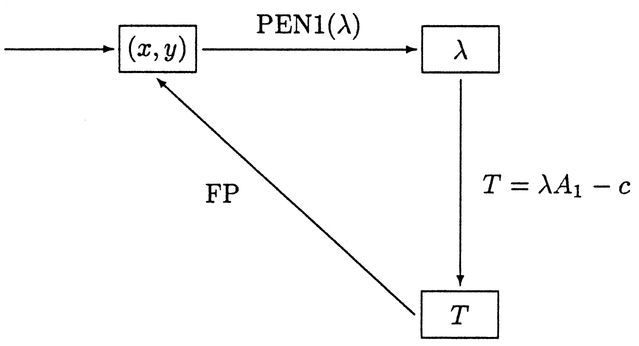

Fig. 2. Heuristic 1: the main steps.

Alternatively, in the spirit of Gendreau, Marcotte, and Savard (1996), one can associate with a dual vector λ an optimal (x, y)-solution to the lower level LP parameterized by the toll vector $T = \lambda A_1 - c$. The strategies that have been retained are described in the next subsections.

2.1 Heuristic 1

Heuristic 1 iterates between the leader's vector T and the follower's vectors x and y. At a given iteration, the tariff vector T is set to $\lambda A_1 - c$, where the dual vector λ solves the penalized problem PEN1 for fixed primal flow vectors x and y and penalty parameter M_1. Next, the flow variables on both the tariff and free arcs correspond to the optimal solution of the lower level distribution problem, with the freight rates of the leader carrier set at $c + T$. So that the penalized problem be bounded, we introduce the artificial constraint,

$$-\lambda^{\max} \leqslant \lambda \leqslant \lambda^{\max}, \tag{11}$$

where $\lambda^{\max}$ is some suitably large constant. (For large values of the penalty parameter, this constraint becomes inactive.) The algorithm is illustrated by Figure 2, where FP designates the follower's problem, and the corresponding pseudo-code. In the diagram, the updating of the penalty parameter is performed at the northeast corner (see steps 5 and 6 of the pseudo-code).

Heuristic 1

Step 1. *Initialization*

1. Set x_0 and y_0 to an optimal solution of the follower's problem with T set to zero.
2. Initialize M_1 and l to 1.
3. Set Z^* (best profit achieved yet) to zero.

Step 2. *Determination of the dual vector* λ

For x_{l-1} and y_{l-1} fixed, solve the problem

PEN1(λ)

$$\max_{\lambda} \; (\lambda A_1 - c)x_{l-1} - M_1(\lambda A_1 x_{l-1} + dy_{l-1} - \lambda b)$$

$$\begin{aligned} \text{subject to} \quad & \lambda A_2 \leqslant d \\ & \lambda \leqslant \lambda^{\max} \\ & -\lambda \leqslant \lambda^{\max}. \end{aligned} \tag{12}$$

Denote its solution by λ_l.

Step 3. *Computation of the tariff vector*

Compute $T_l = \lambda_l A_1 - c$.

Step 4. *Determination of flow vectors x and y*

Let (x_l, y_l) be an optimal solution of the follower's problem

$$\text{FP} \qquad \begin{aligned} \min_{x,y} \; & (c + T_l)x + dy \\ \text{subject to} \; & A_1 x + A_2 y = b \\ & x, y \geqslant 0. \end{aligned}$$

Step 5. *Update of the best profit*

1. Determine the total profit: $Z_l = T_l x_l$,
2. **if** $Z_l > Z^*$ **then** $Z^* \leftarrow Z_l$ and $(T^*, x^*, y^*) \leftarrow (T_l, x_l, y_l)$.

Step 6. *Stopping criterion*

if $|Z_l - Z_{l-1}| < \varepsilon$

then STOP with solution (T^*, x^*, y^*)

else increase M_1, increment l by 1 and go to Step 2.

If, at Step 4 of the algorithm, the solution of the follower's problem is not unique, then one should favor the solution providing the highest profit for the leader, i.e., given a tie, the solution favoring the use of the tariff arcs. This can be achieved by decreasing the tariffs T_a by a small amount ε, which amounts to a local parametric analysis, and requires a few simplex pivot.

At Step 2 of the algorithm, for fixed flow vectors x and y, the penalized problem PEN1(λ) is linear with respect to the dual vector λ. Let us denote by $u \in R^{m_2}$, $z \in R^n$, and $v \in R^n$, the dual variables associated with constraints 12. The dual of problem PEN1(λ) is

$$\min_{u,z,v} \; du + \lambda^{\max} z + \lambda^{\max} v$$

subject to

$$\begin{aligned} & A_2 u + Iz - Iv = (1 - M_1) A_1 x + M_1 b \\ & z, u, v \geqslant 0. \end{aligned} \tag{13}$$

Problem 13 is a transshipment problem on the network $\tilde{G} = (\tilde{\mathcal{N}}, \tilde{\mathcal{A}})$. The node set $\tilde{\mathcal{N}}$ includes nodes of the original network as well as an artificial node f. The arc set $\tilde{\mathcal{A}}$ contains:

- the free arcs $(i, j) \in \mathcal{A}_2$, with costs $d_{i,j}$, whose flow variables are denoted by $u_{i,j}$.
- the arcs linking each node j of $\mathcal{N}$ to the artificial node f, with costs $\lambda^{\max}$, and whose associated flow variables are denoted by z_{jf}.
- the arcs linking the artificial node f to each node i of $\mathcal{N}$, with costs $\lambda^{\max}$, and whose associated flow variables are defined by v_{fj}. These arcs prevent the transshipment problem from being infeasible.

The purpose of the parameter M_1 is to penalize the duality gap $(c + T)x + dy - \lambda b$. However, the vector λ_l obtained at Step 2 of the algorithm is also dual–optimal for problem FP solved at Step 4. Hence, at the end of Step 4, the duality gap is zero. To prove this result, let us write down the optimality conditions of PEN1(λ) where we remove, for simplicity, the bounds on λ, and where the objective,

$$(\lambda A_1 - c)x_{l-1} - M_1(\lambda A_1 x_{l-1} + dy_{l-1} - \lambda b),$$

is replaced by the equivalent scaled objective

$$\lambda\left(b - \left(1 - \frac{1}{M_1}\right) A_1 x_{l-1}\right).$$

The optimality conditions are:

- dual feasibility: $\lambda A_2 \leqslant d$
- primal feasibility: $(1 - 1/M_1)A_1 x_{l-1} + A_2 u = b,\ u \geqslant 0$
- complementarity: $\lambda(b - (1 - 1/M_1)A_1 x_{l-1}) = du$

and we denote by (u_l, λ_l) an optimal solution of the above system. In contrast, the optimality conditions of the linear program FP are:

- primal feasibility: $A_1 x + A_2 y = b,\ x, y \geqslant 0$
- dual feasibility: $\lambda A_1 \leqslant c + T_l = \lambda_l A_1,\ \lambda A_2 \leqslant d$
- complementarity: $\lambda b = \lambda_l A_1 x + dy$.

It is clear that $(x, y, \lambda) = ((1 - 1/M_1)x_{l-1}, u_l, \lambda_l)$ constitutes an optimal (maybe non-basic though) primal–dual solution for the linear program FP, hence λ_l is dual–optimal for FP, and the corresponding duality gap is zero. It follows that, for any optimal primal solution (x^*, y^*) of FP, and in particular (x_l, y_l), the duality gap corresponding to the triple (x^*, y^*, λ_l) is zero.

This result provides some insight on why the number of basis solutions encountered by this algorithm is quite small. Actually, modifications of the

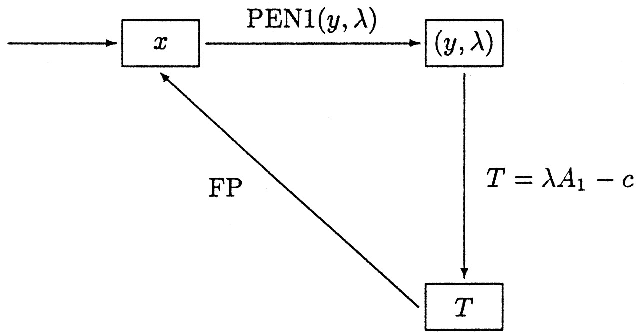

Fig. 3. Heuristic 2: the main steps.

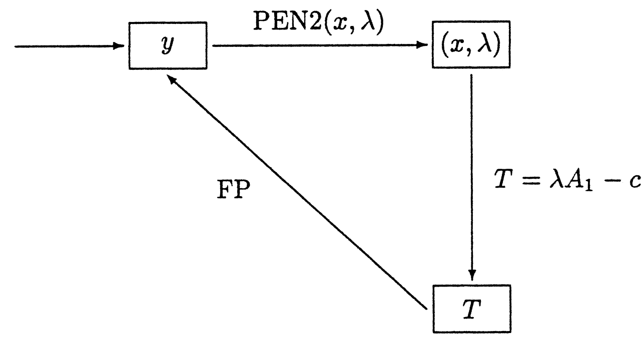

Fig. 4. Heuristic 3: the main steps.

y-vector can only be induced by modifications in the supply–demand patterns resulting from an increase of the penalty parameter M_1. If the corresponding solution $(x, y) = ((1 - 1/M)x_{l-1}, u)$ is not basic, then a change of basis will occur when solving for the follower's problem FP. As a consequence of this remark, it is important to carefully select the initial solution (x^0, y^0). Indeed, if x^0 had been inadvertently set to zero, then it could never have assumed positive values in the sequel, and the profit would have stayed at zero for the entire course of the algorithm! To circumvent that problem, a diversification strategy, to be described in Section 3, has been implemented.

2.2 Heuristic 2 and Heuristic 3

Heuristics 2 and 3 are based on similar principles. The penalized problem PEN1 (respectively PEN2 for heuristic 3) is first solved for fixed values of the vector x (respectively y). The resulting linear program yields the vectors y (x for heuristic 3), λ, and T. In other words, one determines the leader vector T and one follower vector (either x or y) that maximizes total profit while respecting, to some extent, the follower's optimality conditions. Both subproblems are decomposed into a primal and a dual problem.

Next, the vector x (y for heuristic 3) is determined by solving the follower's problem. As in heuristic 1, a constraint is introduced to bound problem PEN1 with respect to the λ variables in heuristic 2. Such a constraint is not required for heuristic 3 since, in PEN2, the variable part of the objective $\lambda b + M_2\lambda A_2 y$ is less than or equal to $M_2 dy + \lambda b$, because $\lambda A_2 \leqslant d$ and flows are nonnegative. Now, due to our assumption that there exists a lower level solution that uses only free arcs, the dual objective λb must be bounded. Heuristics 2 and 3 are illustrated in Figures 3 and 4, from which the pseudo-codes are readily derived along the lines of the pseudo-code of heuristic 1.

As in Section 2.1, the objective of the follower's problem (diagonal arc in the diagrams) is perturbed to induce the optimal lower level solution that maximizes the leader's profit. For fixed flow vector x (y for heuristic 3), the penalized problems PEN1(y, λ), (respectively PEN2(x, λ)) are separable and linear. Exactly as in heuristic 1, one solves their dual problems, which are transshipment problems similar to problem 13.

2.3 Heuristic 4

The last heuristic proposed for the FTSP consists of applying a block Gauss–Seidel procedure to the bilinear problem PEN1 (Fig. 5). More precisely, PEN1 is iteratively solved for fixed flow vectors (x, y) on the one hand, and for the dual vector λ on the other hand. Both problems are linear in λ and (x, y) respectively. As before, constraint 11 is introduced to prevent the occurrence of unbounded solutions in the dual problem, i.e., infeasibility in the transshipment subproblems. Because $T = \lambda A_1 - c$, the objective function 6 of PEN1 can be rewritten as

$$Tx - M_1((c + T)x + dy - \lambda b).$$

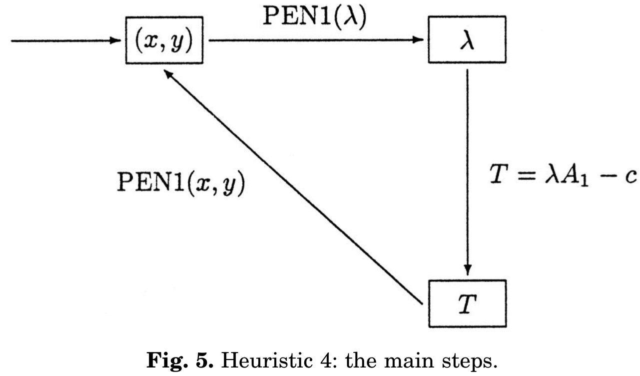

Fig. 5. Heuristic 4: the main steps.

Thus, for fixed x, y vectors, a trade-off is achieved between profit maximization at the upper level and optimality at the lower level. Note that the objective function of the penalized problem PEN1 is monotonically decreasing and has a discontinuity for each new solution (x, y), of the follower's problem solved for a given value of T. Heuristic 4 is described below.

Heuristic 4

Step 1. *Initialization*
1. Set x_0 and y_0 to an optimal solution of the follower's problem with T set to zero.
2. Initialize M_1 to 1 and l to 1.

Step 2. *Determination of the dual vector* λ

For x_{l-1} and y_{l-1} fixed, solve

PEN1(λ)

$$\max_{\lambda} \; (\lambda A_1 - c)x_{l-1} - M_1(\lambda A_1 x_{l-1} + dy_{l-1} - \lambda b)$$

$$\text{subject to} \quad \lambda A_2 \leqslant d$$
$$\lambda \leqslant \lambda^{\max}$$
$$-\lambda \leqslant \lambda^{\max}.$$

The solution is denoted λ_l and we set $T_l = \lambda_l A_1 - c$.

Step 3. *Determination of the flow variables*

For λ_l fixed, solve

PEN1(x, y)

$$\max_{x,y} \; (\lambda_l A_1 - c)x - M_1(\lambda_l A_1 x + dy - \lambda_l b)$$

$$\text{subject to} \quad A_1 x + A_2 y = b$$
$$x, y \geqslant 0.$$

The solutions are denoted x_l, y_l.

Step 4. *Stopping criterion*

If the duality gap $G_l = (c + T_l)x_l + dy_l - \lambda_l b$ is equal to zero

then
1. compute the profit $Z = T_l x_l$ where $T_l = \lambda_l A_1 - c$,
2. STOP with the solution $(T^*, x^*, y^*) \leftarrow (T_l, x_l, y_l)$,

else increase M_1, increment l by 1 and go to Step 2.

Problem PEN1(λ) at Step 2 is solved as in Section 2.1. Note that, for a fixed value of the vector λ, the problem PEN1(x, y) is equivalent to the scaled problem

$$\min_{x,y} \; (c + T_l)x + dy - \frac{1}{M_1} T_l x.$$

This objective can be interpreted as a perturbation of the follower's problem. As M_1 becomes large, the solution coincides with a solution that maximizes the leader's profit while satisfying the optimality conditions of the lower level.

3. IMPROVEMENT STRATEGIES

THE FOUR HEURISTICS developed previously generate a sequence of basic solutions for the lower level problem that may, or may not, correspond to optimal solutions of the original problem FTSP. In this section, we introduce two strategies aimed at generating improved solutions.

3.1 Inverse Optimization

It is frequently the case that the set of arcs that carry positive flow in an optimal or near-optimal solution can be obtained through some (unspecified) heuristic procedure, whereas the exact values of these flow are unknown. If this is the case, one can recover, at small computational cost, the values of these flows that maximize the leader's revenue. This technique was first described in the case of a multicommodity toll problem by Labbé, Marcotte, and Savard (1998).

More specifically, let us consider the situation where lower level vectors x and y are provided, and one wishes to determine a profit-maximizing tariff T that is compatible with the lower level optimality of (x, y). Because the positive entries of the vector y are known, one can get rid of the complementarity constraint in formulation 9 of FTSP, which decomposes into a dual linear program in the variable λ and a primal LP in x and y. Let $\mathcal{A}_+$ denote the set of indices for which y_a is positive. The dual LP takes the form,

$$\max_{\lambda} \; \lambda b$$
$$\text{subject to} \quad (\lambda A_2 - d)_a = 0 \quad \forall a \in \mathcal{A}_+ \qquad (14)$$
$$(\lambda A_2 - d)_a \leqslant 0 \quad \forall a \notin \mathcal{A}_+.$$

The dual of this dual problem is the primal problem

$$\min_{y'} \; dy'$$
$$\text{subject to} \quad A_2 y' = b$$
$$y'_a \text{ unconstrained} \quad \forall a \in \mathcal{A}_+ \qquad (15)$$
$$y'_a \geqslant 0 \quad \forall a \notin \mathcal{A}_+.$$

This is nothing but a transshipment problem on a modified network where the tariff arcs have been deleted and the free arcs carrying positive flow are two-way arcs. The optimal dual vector λ' of this LP yields the desired tax vector $T' = \lambda A_1 - c$. This process, which consists of optimizing an auxiliary objective while forcing some solution to be optimal, is sometimes referred as "inverse optimization."

Note that the sole knowledge of the index set $\mathcal{A}_+$ allows one to conduct the above analysis. However, one would much prefer to base this analysis on the vector x, which will usually be of smaller dimension. Unfortunately, the knowledge of the entire vector x is required to recover the vector y through the solution of the LP:

$$\begin{aligned} &\min_y \; dy \\ &\text{subject to} \quad A_2 y = b - A_1 x \\ &\qquad\qquad\quad y \geq 0. \end{aligned} \tag{16}$$

Of course, knowing the values of the x-variables is equivalent to knowing which ones are positive, whenever the solution of the lower level problem is binary valued. This is not the general situation of a transshipment problem, and we compromized by solving an LP where the positive x-variables are required to be at least one. More specifically, corresponding to an (x, y)-solution with zero duality gap generated by a given heuristic, we solve the linear program,

$$\begin{aligned} &\min_{x,y} \; cx + dy \\ &\text{subject to} \quad A_1 x + A_2 y = b \\ &\qquad\qquad\quad x_a \geq 1 \quad \forall\, a \in \mathcal{A}'_1 \\ &\qquad\qquad\quad x_a = 0 \quad \forall\, a \notin \mathcal{A}'_1, \end{aligned} \tag{17}$$

where $\mathcal{A}'_1$ denotes the set of tariff arcs carrying positive flow. A tariff vector T is then obtained by solving problem 15, with the set $\mathcal{A}_+$ corresponding to the partial optimal solution (in y) of problem 17.

3.2 Diversification

In this section, we consider simple perturbations of the best solution obtained by any given heuristic, in the hope of generating neighboring solutions of higher profit. If the process is successful, then the heuristic is applied starting with this improved solution. In meta-heuristic parlance, this may be interpreted as a diversification strategy around the current best solution. For each tariff arc of the solution with maximum profit, three diversifications are considered. The first one consists of forbidding the use of an arc with positive flow and in initializing the tariffs to zero. The second and third ones consist, respectively, of decreasing by half the tariff associated with an arc or in increasing it by half. Let $\mathcal{AT}^*$ denote the set of tariff arcs at the current best solution (T^*, x^*, y^*), and δ be the index of the selected strategy. The diversification procedure can then be described as follows.

Diversification

Step 1. *Initialization*
1. Let (T^*, x^*, y^*) be the solution produced by a primal–dual heuristic.
2. $Z^* \leftarrow T^* x^*$.
3. $\mathcal{AT}^* \leftarrow \{a \in \mathcal{A}_1 : x^*_a > 0\}$ and $\delta \leftarrow 1$.

Step 2. While $\mathcal{AT}^* \neq \emptyset$, repeat Steps 3 to 6.
Step 3. Let $\tilde{a}$ be the most profitable arc in $\mathcal{AT}^*$.
Step 4. Define the strategy:
If $\delta = 1$ **then** $c_{\tilde{a}} \leftarrow \infty$ and $T_a \leftarrow 0$ for all $a \in \mathcal{A}_1$.
If $\delta = 2$ **then** $T_{\tilde{a}} = \frac{1}{2} T_{\tilde{a}}$.
If $\delta = 3$ **then** $T_{\tilde{a}} = \frac{3}{2} T_{\tilde{a}}$.
Step 5. *Determine a new solution*
1. Solve the follower's problem FP.
2. Apply a primal–dual heuristic starting from this solution.
   Let (T, x, y) the solution and Z the associated profit.

Step 6. *Update of the solution with the best profit*
If $Z > Z^*$ **then**:
$(T^*, x^*, y^*) \leftarrow (T, x, y)$ and $Z^* \leftarrow Z$,
$\mathcal{AT}^* \leftarrow \{a \in \mathcal{A}_1 : x^*_a > 0\}$,
$\delta \leftarrow 1$ and go to Step 2.
else $\delta \leftarrow \delta + 1$.
If $\delta = 4$
then $\delta \leftarrow 1$, $\mathcal{AT}^* \leftarrow \mathcal{AT}^* \backslash \{\tilde{a}\}$ and go to Step 2
else go to Step 3.

4. NUMERICAL RESULTS

THE HEURISTIC procedures developed in this paper have been applied to a set of random instances created using the NETGEN generator of KLINGMAN, NAPIER, and STUTZ (1974). Network sizes range from 50 nodes and 250 arcs to 200 nodes and 9950 arcs. The proportion of tariff arcs varies from 5% to 20%. The arc costs vary from 5 to 35, with 20% of the costs set at their maximum value of 35. In some instances, the tariff arcs are scattered throughout the network, whereas some test problems considered chains of tariff arcs, as would occur in the case of toll highways, for example.

The tariff arcs are generated as follows. Given an ordering of the arcs, the first arc is given the tariff status with probability p if its deletion leaves at least one feasible path from any supply to any demand node of the network. At the kth iteration of the selection process, the kth arc is selected, with probability p, if its deletion, together with that of previously selected tariff arcs, leaves at least one feasible path from any supply to any demand node. The process is halted whenever the required proportion of tariff arcs is reached. The selection process obviously rules out unboundedness of the leader's profit resulting from the presence of captive markets.

To favor the use of tariff arcs by the carriers, we compute the number of times that a given arc is part of a shortest path from a supply to a demand node. According to this ordering, arcs are retained until ⅔ of the total number of desired tariff arcs is attained. The remaining third is selected at random according to the process previously described. Furthermore, to make the tariff arcs attractive, their random costs have been halved.

For generating toll highways, the arcs are sorted with respect to the frequency with which they occur in shortest paths from supply to demand nodes. The procedure first builds the toll highway forward, selecting the arcs in decreasing order of their respective frequencies, until either the list of admissible arcs leaving the current node is empty, a demand node is reached, or the prespecified maximum length of the highway is reached. (Arcs forming a circuit are forbidden.) The process is then performed backward from the current initial node of the path. The entire process is repeated until the number of desired toll highways is reached. If the density of toll arcs is less than required, additional tariff arcs are selected according to the random procedure described in the previous paragraph.

This process is applied to the bipartite network illustrated in Figure 6, where the demand for transportation is set to

$$b = (-1, 1, -1, 0, -1, 2).$$

The shortest paths linking supply nodes to demand nodes are (1, 3, 4, 2), (1, 3, 4, 6), (3, 4, 2), (3, 4, 6), (5, 3, 4, 2), (5, 3, 4, 6). Assume that the length of any toll highway has to be $\leqslant 3$. We first select the arc with the highest frequency, i.e., arc (3, 4) of frequency 6. The next arc to be selected is arc (4, 6), with a frequency of 3. Inasmuch as node 6 is a demand node and the maximal number of arcs of a path is not yet reached, we now proceed backward from the initial node 3. Next, we select arc (5, 3), which is the only remaining acceptable arc. Because node 1 is a demand node, the algorithm stops with path (5, 3, 4, 6).

The transshipment subproblems are solved using the minimum cost flow code of GOLDBERG and TARJAN (1990). To assess the quality of the heuristics, smaller instances of the problems are solved to optimality by feeding the solver CPLEX6.0 (CPLEX, 1993) with the mixed integer programming formulation MIP of the FTSP. The value of M in MIP is fixed to 200 while the penalty factor M_1 is updated according to the parameters given in Table II. The heuristics are coded in C on a Sun Ultra 60 (360 Mhz).

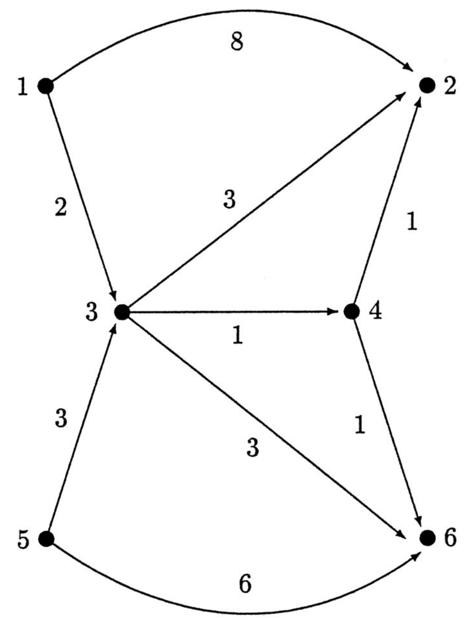

Fig. 6. Definition of a tariff path.

The numerical results are summarized in Table III. The reports for Heuristics 1, 2, and 3 involve a diversification phase. This strategy having proved too costly for heuristic 4, the corresponding results are not reported. The first column (%T) provides the percentage of tariff arcs. The second column (SN) gives the number of supply nodes (the number of demand nodes is set equal to the number of supply nodes). The third column (NB) indicates the number of instances where CPLEX has been halted, either because the number of nodes or the CPU times exceeded the upper limit of 400,000 nodes or 18,000 seconds.

Each algorithm has been tested on a series of 10 randomly generated problems. For each algorithm

TABLE II
Update of the Penalty Factor M_1

	H1	H2	H3	H4
initial value	1	1.01	1.1	1.1
increment	1	0.01	0.01	0.01

TABLE III
Numerical Results

%T	SN	NB	H1		H1DIV		H2		H2DIV		H3		H3DIV		H4		CPLEX	
Networks with 50 nodes and 250 links																		
5	5	0	0.968	0.57	1.000	0.92	0.917	0.12	0.990	0.47	0.960	0.13	0.996	0.54	1.000	0.23	25863	62
			0.090	0.43	0.001	0.24	0.147	0.02	0.028	0.06	0.090	0.01	0.008	0.11	0.001	0.07	41065	101
10	5	2	0.971	0.15	0.998	1.11	0.916	0.13	0.996	0.82	0.984	0.14	0.998	0.66	0.990	0.25	17866	143
			0.072	0.02	0.005	0.15	0.144	0.01	0.007	0.19	0.038	0.01	0.005	0.07	0.023	0.10	31873	204
15	5	1	1.000	0.14	1.000	1.08	0.995	0.12	0.995	0.73	0.999	0.12	1.000	0.82	1.000	0.21	11785	111
			0.001	0.02	0.000	0.35	0.014	0.02	0.015	0.17	0.002	0.02	0.000	0.20	0.000	0.13	19802	186
20	5	2	0.999	0.14	1.000	1.23	1.000	0.11	1.000	0.64	0.999	0.12	1.000	0.63	1.000	0.11	16953	181
			0.002	0.01	0.000	0.23	0.000	0.02	0.000	0.10	0.002	0.01	0.000	0.09	0.000	0.02	27067	237
Networks with 75 nodes and 1400 links																		
5	10	4	0.995	0.83	0.996	7.28	0.988	0.57	0.988	5.29	0.989	0.52	0.997	3.89	0.997	1.72	51230	1563
			0.006	0.29	0.006	2.26	0.016	0.09	0.016	1.46	0.016	0.07	0.006	1.04	0.006	0.87	49136	1739
10	10	3	0.999	0.63	0.999	8.87	0.999	0.51	0.999	4.79	0.999	0.49	0.999	4.19	1.000	0.72	12789	1082
			0.002	0.02	0.002	0.83	0.002	0.04	0.002	1.34	0.002	0.03	0.002	0.31	0.000	0.23	9933	911
15	10	2	1.000	0.65	1.000	9.84	1.000	0.52	1.000	5.64	0.997	0.48	1.000	4.55	1.000	0.86	26770	2475
			0.000	0.01	0.000	0.98	0.001	0.04	0.001	1.04	0.008	0.02	0.000	0.74	0.000	0.44	36877	246
20	10	0	1.000	0.60	1.000	9.43	1.000	0.50	1.000	5.77	1.000	0.49	1.000	4.90	1.000	0.78	28668	6051
			0.000	0.03	0.000	2.33	0.000	0.04	0.000	1.54	0.000	0.03	0.000	1.41	0.000	0.14	33021	6123
Networks with highways (50 nodes and 250 links or 75 nodes and 1400 links)																		
N1-10	5	1	0.996	0.14	0.997	1.16	0.971	0.14	0.984	0.85	0.996	0.13	0.997	0.99	1.000	0.35	2553	55
			0.010	0.01	0.001	0.28	0.061	0.01	0.029	0.16	0.010	0.02	0.010	0.40	0.000	0.18	1876	110
N1-20	5	0	1.000	0.13	1.000	1.34	1.000	0.13	1.000	0.78	1.000	0.12	1.000	0.73	1.000	0.13	3097	40
			0.000	0.02	0.000	0.21	0.000	0.02	0.000	0.21	0.000	0.01	0.000	0.10	0.000	0.01	800	12
N2-10	10	0	1.000	0.65	1.000	8.56	1.000	0.52	1.000	5.01	1.000	0.51	1.000	4.70	1.000	0.94	6945	547
			0.000	0.02	0.000	1.12	0.000	0.02	0.000	0.72	0.000	0.02	0.000	0.73	0.000	0.06	4628	444
N2-20	10	0	1.000	0.63	1.000	8.89	1.000	0.50	1.000	4.96	1.000	0.50	1.000	4.77	1.000	0.93	8883	1533
			0.000	0.01	0.000	0.82	0.000	0.02	0.000	0.60	0.000	0.02	0.000	0.45	0.000	0.04	7038	1795
Networks with 100 nodes and 2475 links																		
1	15		0.964	1.70	0.99	10.63	0.834	1.50	0.872	6.93	0.947	1.70	0.98	8.04	0.996	14.67		
			0.035	0.15	0.022	4.12	0.093	0.06	0.099	2.22	0.058	0.17	0.022	1.64	0.007	8.24		
2.5	15		0.986	1.40	0.995	18.17	0.971	1.30	0.972	11.67	0.983	1.40	0.994	12.19	0.996	10.41		
			0.015	0.21	0.010	5.22	0.022	0.12	0.022	2.89	0.020	0.12	0.012	3.81	0.007	4.86		
5	15		0.999	1.30	0.999	22.40	0.998	1.10	0.998	14.05	0.999	1.00	0.999	11.82	1.000	2.62		
			0.002	0.10	0.002	3.68	0.003	0.07	0.003	2.10	0.002	0.12	0.002	2.31	0.000	1.87		
10	15		1.000	1.3	1.000	24.23	1.000	1.06	1.000	12.9	1.000	1.06	1.000	11.93	1.000	1.52		
			0.000	0.12	0.000	3.78	0.000	0.09	0.000	2.21	0.000	0.09	0.000	2.27	0.000	0.91		
Networks with 200 nodes and 9950 links																		
1	30		0.979	9.25	0.984	179.01	0.924	8.32	0.945	120.18	0.980	8.91	0.987	107.86	0.998	165.20		
			0.021	0.55	0.020	51.37	0.061	0.42	0.029	41.18	0.021	0.84	0.011	18.80	0.006	94.40		
2.5	30		0.992	9.77	0.997	242.84	0.995	8.69	0.996	142.47	0.993	8.99	0.996	145.24	1.000	92.39		
			0.012	0.58	0.006	55.43	0.006	0.50	0.006	27.83	0.013	0.85	0.006	18.13	0.001	35.11		
5	30		0.996	8.11	1.000	279.51	0.999	7.96	0.999	157.37	0.999	7.94	1.000	166.89	0.999	75.23		
			0.005	0.53	0.000	55.35	0.001	0.78	0.001	20.00	0.002	1.02	0.000	48.71	0.002	35.92		

(with the exception of CPLEX), the first line displays the average ratio of the heuristic objective over the optimal value, and the CPU time (in seconds). On the second line of each cell, the corresponding standard deviations are displayed. In the first three sections, the averages and standard deviations have been taken only over those problems that could be solved to optimality (10 − NB).

In the first CPLEX column, the optimality ratios have been replaced by the number of nodes explored in the branch-and-bound tree.

On small networks, the heuristics are much faster than CPLEX. Typically, the number of iterations required by the primal–dual methods 1, 2, or 3 (without diversification) is less than four, whereas it varies between 1 and 26 for the fourth method. On the large instances, heuristics 1, 2, and 3 (without diversification) require less than 10 seconds. This number increases to 280 seconds if the diversification phase is implemented. CPU times vary from 2 to 167 seconds for heuristic 4. We note that the CPU time for heuristic 4 decreases when the percentage of tariff arcs increases. In all cases, the number of iterations is quite insensitive to the size of the network.

Heuristics 2, 3, and 4 are halted as soon as a solution with zero duality gap is obtained. This strategy was justified by preliminary experiments that revealed that this solution is seldom improved upon. Note also that distinct optimal solutions, i.e., solutions involving distinct tariff vectors, are sometimes obtained by different heuristics.

As a general rule, heuristics 1, 2, and 3 with the diversification phase, and heuristic 4, based on the Gauss–Seidel method, sharply outperform the other heuristic procedures. On the smaller instances, the primal–dual heuristics 1 and 3 without the diversification phase produce, on the average, solutions within 0.7% of optimality (the worst case being at 4%), whereas 1 and 3 with the diversification phase, and 4, fall within 0.3% of optimality (the worst case at 0.4%) on the average. On larger test problems, heuristic 4 comes up slightly better than either 1 or 3 with the diversification strategy, the solutions being within 1% of the best known solution, which is quite good.

5. CONCLUSION

NOT ONLY ARE tarification problems pervasive in decision making, but they also constitute a rich class of structured bilevel problems. In this paper, we considered a special member of this class, which is amenable to solution techniques that allow us to solve instances of significant size within reasonable computing times.

Research is currently underway on a multi-commodity version of this problem, where the lower level consists of individuals traveling on arcs of a network subject to tolls set by a profit-maximizing leader.

REFERENCES

G. ANANDALINGAM AND D. J. WHITE, "A Solution Method for the Linear Stackelberg Problem Using Penalty Functions," *IEEE Trans. Autom. Control* **35,** 1170–1173 (1990).

C. AUDET, P. HANSEN, B. JAUMARD, AND G. SAVARD, "A Symmetrical Linear Maxmin Approach to Disjoint Bilinear Programming," *Math. Program.* **85,** 573–592 (1999).

L. BROTCORNE, Approches opérationnelles et stratégiques des Problèmes de trafic routier, Ph.D. thesis, Université Libre de Bruxelles, Belgium (1998).

CPLEX OPTIMIZATION INC., Using the CPLEX Callable Library and CPLEX Mixed Integer Library, 1993.

C. S. FISK, "A Conceptual Framework for Optimal Transportation Systems Planning with Integrated Supply and Demand Models," *Transp. Sci.* **20,** 37–47 (1986).

T. L. FRIESZ, J. A. GOTTFRIED, AND E. K. MORLOK, "A Sequential Shipper–Carrier Network Model for Predicting Freight Flows," *Transp. Sci.* **20,** 80–91 (1986).

M. GENDREAU, P. MARCOTTE, AND G. SAVARD, "A Hybrid Tabu Ascent Algorithm for the Linear Bilevel Programming Problem," *J. Global Optimiz.* **8,** 217–233 (1996).

A. V. GOLDBERG AND R. E. TARJAN, "Finding Minimum-Cost Circulation by Successive Approximation," *Math. Opns. Res.* **15,** 430–466 (1990).

P. T. HARKER, "Multiple Equilibrium Behaviors on Networks," *Transp. Sci.* **22,** 39–46 (1988).

W. HURLEY AND E. R. PETERSEN, "Optimal Freight Transport Pricing and the Freight Network Equilibrium Problem," Proc. of TRISTAN II, Capri, 1994a, 255–265.

W. HURLEY AND E. R. PETERSEN, "Nonlinear Tariffs and Freight Network Equilibrium," *Transp. Sci.* **28,** 236–245 (1994b).

D. KLINGMAN, A. NAPIER, AND J. STUTZ, "Netgen: A Program for Generating Large Scale Capacitated Assignment, Transportation, and Minimum Cost Flow Network Problems," *Management Sci.* **20,** 814–821 (1974).

M. LABBÉ, P. MARCOTTE, AND G. SAVARD, "A Bilevel Model of Taxation and Its Application to Optimal Highway Pricing," *Management Sci.* **44,** 1595–1607 (1998a).

M. LABBÉ, P. MARCOTTE, AND G. SAVARD, "On a Class of Bilevel Programs," Forthcoming in the Proceedings of the 26th Workshop: Nonlinear Optimization and Applications, International School of Mathematics, G. Stampacchia, Erice, Italy, 1998b.

P. MARCOTTE, "Algorithms for the Network Oligopoly Problem," *J. Oper. Res. Soc.* **38,** 1051–1065 (1987).

P. REY AND J. TIROLE, "The Logic and Vertical Restraint," *Amer. Economic Rev.* **76,** 921–939 (1986).

T. V. THIEU, "A Note on the Solution of Bilinear Problems by Reduction to Concave Minimization," *Math. Program.* **41,** 249–260 (1988).

J. TIROLE, *A Theory of Industrial Organisation*, MIT Press, 1989.

R. WILSON, *Nonlinear Pricing*, Oxford University Press, New York, 1993.

(Received: February 1999; revisions received: November 1999; accepted: November 1999)