

## Finding second-best toll locations and levels by relaxing the set of first-best feasible toll vectors

Joakim Ekström<sup>1</sup>

Linköping University, Norrköping, Sweden

---

This paper provides a framework for optimizing toll locations and levels in congestion pricing schemes for large urban road networks, with the objective to maximize the social surplus. This optimization problem is referred to as the toll location and level setting problem (TLLP) and is both non-convex, non-smooth and involves binary decision variables, and is therefore considered as a hard problem to solve. In this paper a solution approach is provided which instead of directly solving the TLLP, makes use of the first-best toll level solution, in which no restrictions are imposed on toll locations or levels. A first-best pricing scheme can be obtained by solving a convex program, and it has previously been shown that for the used routes in the network, the first-best toll levels on a route level are unique. By formulating an optimization problem, which instead of maximizing the social surplus, tries to find the link toll levels which minimize the deviation from first-best route tolls, a mixed integer linear program is obtained, and if the toll locations are predetermined the resulting optimization problem is a linear program.

The approach of minimizing the deviation from first-best route tolls is applied for two different network models, and results are provided to show the applicability of the approach, as well as to compare with other approaches. Also, it is shown that for the Stockholm network, virtually the first-best level of social surplus can be obtained with a significantly reduced number of located tolls.

*Keywords:* congestion pricing; network design; user equilibrium

---

### 1. Introduction

Marginal social cost pricing (MSCP) in a road traffic network will lead to the most efficient usage of the road infrastructure (first discussed by Pigou, 1920), and will maximize the social surplus. One drawback with MSCP is, however, that in practice, a toll needs to be collected on every road segment. Even though collecting a toll on every road segment is possible with the technology available today, it would result in a pricing scheme in which it would be virtually impossible for the road users to predict the cost associated with choosing a specific route through the network. Also, such a pricing scheme may become expensive to implement and operate. Therefore congestion pricing schemes are usually implemented, in practice, with a more limited number of toll locations, and in such a way that the road users easily can understand the system and predict their costs of travelling. The pricing scheme can be in the form of cordon pricing, where the road users pay a toll when crossing a cordon, in the form of area pricing, where the road users pay a toll for accessing a restricted part of the city, or as in the form of distance based pricing where the road users are charged a fixed amount per kilometer driven. In all of these variants of pricing schemes, restrictions are imposed on toll locations and/or toll levels, but the aim is still to

---

<sup>1</sup> A: Campus Norrköping, SE-601 74 Norrköping, Sweden T:+4611265850 F:+4611363011 E: joakim.ekstrom@liu.se

maximize the social surplus and such pricing principles are commonly referred to as second-best ones, as opposed to first-best ones in which there are no restrictions.

When predicting effects of changes in transportation infrastructure or pricing of using the infrastructure it is commonly assumed that the travelers are distributed according to a Wardropian user equilibrium. In a user equilibrium no traveler can reduce his/her travel cost by altering the choice of mode or route in the transportation network. While a distribution of road users according to a user equilibrium assumes that the road users have perfect information about travel costs (both time and monetary costs) within the road network, and that they make decisions which maximize their individual utility, the user equilibrium has desirable mathematical properties. Assuming that the traffic conditions are static over the studied time period, the problem of determining the user equilibrium can be formulated as a convex program. The problem of finding optimal toll locations and toll levels, which maximize the social surplus, can then be formulated as a bi-level optimization program. In which the lower level program is a convex user equilibrium problem, and on the upper level, toll locations and levels are adjusted to maximize the social surplus. A special case of the bi-level program arise when there are no restrictions on the toll locations and their toll levels (first-best pricing), and for this case the problem is convex. For the first-best case, MSCP tolls always give an optimal solution but there can also be alternative optimal solutions. The bi-level program, which is both non-convex and non-smooth for the general case, is similar to other bi-level programs arising in transportation network design problems. For a review on bi-level programs within transportation planning see Migdalas (1995), and for a more recent review on models and methods the case of congestion pricing see Tsekeris and Voß (2009).

For the case when the toll locations are predetermined, the problem is reduced to finding optimal toll levels, which is still a non-convex and non-smooth problem, due to the intricate relationship between the upper and lower level problems. This will be referred to the toll level setting problem (TLP), and it has previously been solved with both ascent methods (Yang and Lam, 1996; Verhoef, 2002a; Chen and Bernstein, 2004; Lawphongpanich and Hearn, 2004; Ekström et al., 2009) and meta-heuristic approaches (Yin, 2000; Yang and Zhang, 2003; Shepherd and Sumalee, 2004; Zhang and Yang, 2004). Introducing stochastic route choices, ascent methods are developed in (Chen et al., 2004; Ying and Yang, 2005; Sumalee et al., 2006; Connors et al., 2007). While the ascent approaches can only be used for finding a local optimal solution and need to deal with the non-differentiability of the TLP, meta-heuristic approaches usually require a large number of toll level solutions to be evaluated, and each evaluation of a solution to the TLP requires one user equilibrium problem to be solved. On the other hand, meta-heuristic approaches include mechanism to avoid getting stuck in local optimal solutions, although they can neither guarantee local or global optimality of the solution.

The toll location and level setting problem (TLLP) has almost exclusively been studied with different heuristic approaches (Verhoef, 2002b; Zhang and Yang, 2004; Sumalee, 2004; Shepherd and Sumalee, 2004; Ekström et al., 2009). Heuristic methods are, however, often based on hierarchal decisions of toll locations and toll levels, and to evaluate one specific toll location require one TLP to be solved, and to achieve good results with such heuristics require a large number of toll locations to be evaluated. The approach in Ekström et al. (2009) is instead based on a smoothening of the discrete part of the objective function, in order to simultaneously determine the optimal number of tolls to locate and the corresponding toll locations and toll levels, based on a cost associated with each toll location, and this approach has been demonstrated on a network model of Stockholm in Ekström et al. (2014).

More recently, global optimization approaches have been adopted to solve the TLLP by Zhang and van Wee (2012) and Ekström et al. (2012). These approaches are based on piecewise linearization of the non-linear functions in the TLLP, resulting in a mixed integer linear program (MILP). So far, global optimization approaches have, however, only been demonstrated on small network models, due to the computational burden of solving the resulting MILPs.

Based on first-best pricing schemes Yildirim and Hearn (2005) formulate a set of valid toll vectors, which fulfil the requirement of being first-best pricing schemes. Different objective functions can then be formulated to choose between the different first-best toll vectors, e.g. the toll vector with minimum number of located tolls or which minimize the maximum toll level. Larsson and Patriksson (1998) and Yin and Lawphongpanich (2009) have, however, shown that while the toll levels differ between different first-best pricing schemes on a link-level, the sum of the collected tolls for each used route is equal for all first-best pricing schemes, under the assumption that the traffic demand is depending on the cost of travelling. Thus, as Larsson and Patriksson (1998) points out, the set of first-best toll vectors is of limited interest as the number of tollable links can only be reduced if several links can be replaced by a single one.

The study in Ekström et al. (2014) suggests that it is, for the presented Stockholm study, possible to achieve 96% of the social surplus associated with first-best pricing, with only 24% of the links being tolled. To search for first-best pricing schemes will result in pricing schemes which account for 100% of the social surplus, but will on the other hand require an increased number of links to be tolled. Thus, close to first-best pricing may be possible using a significantly reduced number of tolled links. In the work presented in this paper, the maximization of the social surplus is replaced by the minimization of the deviation from first-best pricing on a route level. The resulting optimization problem is a linear program (LP), if the toll locations are fixed and a MILP if the toll locations are variable. While the minimization of the deviation from first-best route tolls is not likely to result in toll levels which maximize the social surplus function, it is shown in the numerical results that good toll levels and locations can be obtained within a reasonable computational time for a network model of Stockholm.

The main contribution of this paper is to provide a framework for optimizing toll locations and toll levels, which is applicable to large urban road networks within reasonable computational time. In the numerical results it is shown that although the deviation from the first-best solution is large, toll level solutions which minimize this deviation results in toll levels close to what has been obtained with other, more computational demanding, methods. This paper also presented results which show that congestion pricing schemes resulting in a level of social surplus close to what is reached with first-best pricing can be achieved with a significantly reduced number of located tolls. While Larsson and Patriksson (1998) points out that the set of first-best toll vectors is of limited interest, the results presented in this paper shows that for a practical case, a small relaxation of the first-best toll set can significantly reduce the number of tolled links with virtually the same level of social surplus.

The resulting LP (for the case of fixed toll locations) includes constraints which rely on the complete set of routes being explicitly formulated for every OD-pair. In the optimum solution it is, however, assumed that only a subset of these constraints are binding. Therefore a model, with only a subset of constraints included, is formulated, and an iterative solution algorithm is developed for generating additional routes. For the case of variable toll locations, the resulting MILP is solved with a greedy heuristic.

The remainder of the paper is outlined as follows. In Section 2 the set of first-best toll vectors is formulated, and later relaxed in Section 3 to allow for second-best solutions. Together with the relaxed set of first-best toll vectors an optimization problem is formulated in order to minimize the deviation from first-best route tolls. To solve the resulting LP and MILP, solution algorithms are developed in Section 4 for both the cases of fixed and variable toll locations. Numerical results are presented in Section 5 for network models of Sioux Falls and of Stockholm, and in Section 6 the results are summarized and further research directions are discussed.

## 2. Valid toll vectors in the first-best pricing problem

Consider a road traffic network with a set of origin-destination (OD) pairs  $I$  and a set of links  $A$ . For each OD-pair  $i \in I$  there is a set of routes  $\Pi_i$ , each route  $p \in \Pi_i$  with  $f_p$  being the number of travelers using route  $p$  per hour. The number of cars per hour,  $v_a$ , on link  $a$  is given by  $v_a = \sum_{i \in I} \sum_{p \in \Pi_i} f_p \delta_p^a$  where  $\delta_p^a$  takes on the value of 1 if route  $p$  traverses link  $a$ , and 0 otherwise. The link travel cost is given by  $c_a(\tau_a, v_a) = \alpha t_a(v_a) + \tau_a$ , where  $t_a(v_a)$  is the travel time (in minutes) on link  $a$  at flow  $v_a$ ,  $\tau_a$  the toll level on link  $a$ , and  $\alpha$  the value of time which transforms time into the same monetary unit as the toll levels are given in. The relationship between travel cost and demand is expressed by the inverse travel demand function, which for OD-pair  $i \in I$  is given by  $\pi_i = D_i^{-1}(q_i)$ , where  $\pi_i$  is the minimum travel cost and  $q_i$  the travel demand in OD-pair  $i$  in the unit of travelers per hour. Note that the link flow is given in the unit of cars per hour, and the demand and route flows in the unit of travelers per hour, thus the mean car occupancy,  $\chi$ , will be used to provide a conversion between travelers and cars. Throughout this paper the travel time functions are assumed to be separable and increasing functions and the inverse travel demand functions are assumed to be separable and decreasing functions.

The problem of finding the user equilibrium link flow and demand distribution, given a toll vector  $\bar{\tau}$ , can then be formulated as the complementarity problem (Sheffi, 1985)

$$f_p > 0 \Rightarrow \sum_{a \in A} c_a(\bar{\tau}, v_a) \delta_p^a = \pi_i, \quad p \in \Pi_i, i \in I \quad (1a)$$

$$f_p = 0 \Rightarrow \sum_{a \in A} c_a(\bar{\tau}, v_a) \delta_p^a \geq \pi_i, \quad p \in \Pi_i, i \in I \quad (1b)$$

$$q_i > 0 \Rightarrow D_i^{-1}(q_i) = \pi_i, \quad i \in I \quad (1c)$$

$$q_i = 0 \Rightarrow D_i^{-1}(q_i) \geq \pi_i, \quad i \in I \quad (1d)$$

$$f_p \geq 0, \quad p \in \Pi_i, i \in I \quad (1e)$$

$$q_i \geq 0 \quad i \in I \quad (1f)$$

$$v_a - \frac{1}{\chi} \sum_{i \in I} \sum_{p \in \Pi_i} f_p \delta_p^a = 0, \quad a \in A. \quad (1g)$$

$$\sum_{p \in \Pi_i} f_p = q_i, \quad i \in I \quad (1h)$$

Constraints (1a) and (1b) states that any used route will have a cost equal to the minimum cost,  $\pi_i$ , of making a trip in OD-pair  $i \in I$ , and any unused route will have a cost equal to or larger than  $\pi_i$  for OD-pair  $i \in I$ . For each OD-pair, constraints (1c) and (1d) will ensure that a potential road user only makes a trip if the individual surplus with making the trip exceeds the minimum cost of travelling in the OD-pair. Constraints (1e) and (1f) ensure non-negative route flows and demands respectively, (1g) gives the conversion between route and link flows, and (1h) ensure that the sum of the route flows in each OD-pair equals the demand. For a given toll vector,  $\bar{\tau}$ , the user equilibrium link flows and demands can be obtained by solving the following convex program (Sheffi, 1985)

$$\min_{q, v} G(\bar{\tau}, q, v) = \chi \sum_{a \in A} \int_0^{v_a} c_a(\bar{\tau}_a, u) du - \sum_{i \in I} \int_0^{q_i} D_i^{-1}(w) dw \quad (2)$$

subject to constraints (1e)-(1h).

The efficiency of a congestion pricing scheme is evaluated by the change in social surplus. The social surplus measure ( $SS$ ), for a given toll vector  $\bar{\tau}$  consists of the consumer surplus ( $CS$ ) plus the operator surplus ( $OS$ ). The consumer surplus is expressed as

$$CS(\tau, q(\tau), v(\tau)) = \sum_{i \in I} \int_0^{q_i(\tau)} D_i^{-1}(w) dw - \chi \sum_{a \in A} \left( \alpha t_a(v_a(\tau)) + \frac{\tau_a}{\chi} \right) v_a(\tau), \quad (3)$$

in which the first sum is the user benefits (given by the Marshallian measure (Zerbe and Dively, 1994) and the second sum is the user costs. The operator surplus ( $OS$ ) is equal to the collected tolls

$$OS(\tau, q(\tau), v(\tau)) = \sum_{a \in A} \tau_a v_a(\tau). \quad (4)$$

The social surplus measure can then be expressed as

$$\begin{aligned} SS(\tau, q(\tau), v(\tau)) &= CS(\tau, q(\tau), v(\tau)) + OS(\tau, q(\tau), v(\tau)) \\ &= \sum_{i \in I} \int_0^{q_i(\tau)} D_i^{-1}(w) dw - \chi \sum_{a \in A} \alpha t_a(v_a(\tau)) v_a(\tau). \end{aligned} \quad (5)$$

Let  $v^0$  and  $q^0$  bet the link flow and demand vectors corresponding with the non-tolled solution. The social surplus for the non-tolled solution is then given by

$$SS(q^0, v^0) = \sum_{i \in I} \int_0^{q_i^0} D_i^{-1}(w) dw - \chi \sum_{a \in A} \alpha t_a(v_a^0) v_a^0,$$

with the change in social surplus,  $\Delta SS(\tau)$ , induced by the toll vector  $\tau$ , given by

$$\Delta SS(\tau) = SS(\tau, q(\tau), v(\tau)) - SS(q^0, v^0). \quad (6)$$

Consider the problem of maximizing the social surplus by adjusting the toll locations and toll levels. Let  $y_a$  be a variable for each link  $a \in A$  which takes on the value of 1 if link  $a$  is tolled and 0 otherwise. The feasible combinations of toll locations and levels are then given by the set

$$T^V : \left\{ (\tau_a, y_a) \mid y_a \leq g_a, 0 \leq \tau_a \leq \tau_a^U y_a, a \in A, \sum_{a \in A} y_a = k \right\} \quad (7)$$

where  $k$  is the number of tolled links,  $g_a$  is a parameter which takes on the value of 1 if link  $a$  is tollable and 0 otherwise and  $\tau_a^U$  is the maximum toll level allowed to be charged on link  $a$ . For the case when the toll locations are fixed, the feasible set of toll levels can be formulated as

$$T^F : \left\{ \tau_a \mid 0 \leq \tau_a \leq \tau_a^U g_a \right\} \quad (8)$$

The objective in both the TLP and TLLP is to maximize (6), and since  $SS(q^0, v^0)$  is constant, (5) can be used as objective function. The TLLP can then be formulated as

$$\max_{(\tau, y) \in T^V} F(\tau) = SS(\tau, q(\tau), v(\tau)) = \sum_{i \in I} \int_0^{q_i(\tau)} D_i^{-1}(w) dw - \chi \sum_{a \in A} \alpha t_a(v_a(\tau)) v_a(\tau). \quad (9)$$

For fixed toll locations,  $(\tau, y) \in T^V$  is simply replaced by  $\tau \in T^F$  in (9). Note that  $q(\tau)$  and  $v(\tau)$  are given implicitly by the solution to the user equilibrium problem. This implicit relation between the objective function and toll levels give rise to the non-convex nature of the bi-level program. If  $g_a = 1$  for every link  $a$ , and  $k$  is equal to the number of links in the network, (9) is a convex program with the MSCP toll vector as one, among possibly several, first-best optimal

solutions. The corresponding link flow and demand solution is for this case commonly referred to as the system optimal (SO) link flow and demand solution, and any first-best toll level solution results in SO link flows and demands.

Let  $\pi^{\text{SO}}$ ,  $v^{\text{SO}}$  and  $f_p^{\text{SO}}$  be the vectors of minimum OD travel costs, link flows and route flows respectively, corresponding to the SO link flow and demand solution. Inserting  $\pi^{\text{SO}}$ ,  $v^{\text{SO}}$  and  $f_p^{\text{SO}}$  in (1), Yildirim (2001) shows that any toll vector  $\tau$  satisfying

$$f_p^{\text{SO}} > 0 \Rightarrow \sum_{a \in A} (\alpha t_a(v_a^{\text{SO}}) + \tau_a) \delta_p^a = \pi_i^{\text{SO}}, \quad p \in \Pi_i, i \in I \quad (10a)$$

$$f_p^{\text{SO}} = 0 \Rightarrow \sum_{a \in A} (\alpha t_a(v_a^{\text{SO}}) + \tau_a) \delta_p^a \geq \pi_i^{\text{SO}}, \quad p \in \Pi_i, i \in I, \quad (10b)$$

will maximize the social surplus. Note that in order to formulate the set of first-best toll vectors (10), it is actually only necessary to know the SO link travel times, minimum route travel cost and the set of used routes, which can be computed from  $v^{\text{SO}}$  and  $q^{\text{SO}}$ .

The set of route flows and demands satisfying (10) is denoted as the set of feasible first-best toll vectors. Larsson and Patriksson (1998) show that if the demand is elastic, the total paid toll for each route through the network with a positive flow is independent of the choice of toll vector, as long as the toll vector belong to the set of first-best toll vectors. In Hearn and Ramana (1998) and Yildirim (2001) an equivalent link based formulation of (10) is used to maximize or minimize different objectives over the set of feasible first-best toll vectors, e.g. to minimize the number of located tolls or the minimum of the maximum toll level. The resulting optimization problem is either a LP or a MILP. First-best solutions are, however, usually not possible in practice, when there are restrictions on toll locations and/or toll levels. In the next section the set of feasible toll vectors (10) is therefore relaxed, to allow for second-best solutions.

### 3. Relaxing the first-best toll set

Let  $\Pi_i^{1,\text{SO}}$  and  $\Pi_i^{2,\text{SO}}$  be the set of routes in OD-pair  $i$  with flow  $f_p > 0$  and with flow  $f_p = 0$  respectively, i.e. the set of used and unused routes respectively. The set of first-best feasible toll vectors (10) can then be expressed as

$$\sum_{a \in A} (\alpha t_a(v_a^{\text{SO}}) + \tau_a) \delta_p^a = \pi_i^{\text{SO}}, \quad p \in \Pi_i^{1,\text{SO}}, i \in I \quad (11a)$$

$$\sum_{a \in A} (\alpha t_a(v_a^{\text{SO}}) + \tau_a) \delta_p^a \geq \pi_i^{\text{SO}}, \quad p \in \Pi_i^{2,\text{SO}}, i \in I, \quad (11b)$$

Note that the first-best toll, charged for travelling on a route  $p \in \Pi_i$  in OD-pair  $i \in I$ , can be expressed as  $\tau_p^{\text{route}} = \pi_i^{\text{SO}} - \sum_{a \in A} \alpha t_a^{\text{SO}} \delta_p^a$ . Thus, instead of charging a toll on each link, an equivalent route toll,  $\tau_p^{\text{route}}$ , can be charged to each user on route  $p$ . While there may exist several link toll vectors which are valid in (11), Larsson and Patriksson (1998) show that, under the assumption of elastic demand, the route tolls can be uniquely determined, and this route toll will always be equal to the sum of the first-best link tolls along route  $p$ . For a route with zero flow, it is only necessary for the route toll to be equal to or exceed  $\pi_i^{\text{SO}}$ . For unused routes,  $\tau_p^{\text{route}}$  can therefore be equal to zero if the route cost, even without tolls, exceeds  $\pi_i^{\text{SO}}$ .

The problem of maximizing the social surplus (5) over the set of feasible toll vectors is a non-linear program, if the set of feasible toll vectors are defined by (8), or a mixed integer non-linear program, if the set of feasible toll vectors are defined by (7). Both of these problems belong to the

class of mathematical programs with equilibrium constraints (MPEC) and are in the general case non-convex and non-smooth problems. Instead of directly solving the non-linear program or the mixed integer non-linear program, this paper proposes another strategy for obtaining toll levels and locations in order to maximize the social surplus. By searching for link toll vectors which minimize the deviation first-best route tolls, toll levels can be obtained by solving a LP and toll locations and levels by solving a MILP.

The set (11) can be relaxed by introducing the variables  $\rho_p \geq 0$  and  $\sigma_p \geq 0$  into (11a) for each route  $p \in \Pi_i^{1,SO}, i \in I$ , and the variables  $\mu_i \geq 0$  into (11b) for each OD-pair  $i \in I$ . The relaxed set of first-best toll vectors can then be expressed as

$$\sum_{a \in A} (\alpha t_a(v_a^{SO}) + \tau_a) \delta_p^a + \rho_p - \sigma_p = \pi_i^{SO}, \quad p \in \Pi_i^{1,SO}, i \in I \quad (12a)$$

$$\sum_{a \in A} (\alpha t_a(v_a^{SO}) + \tau_a) \delta_p^a + \mu_i \geq \pi_i^{SO}, \quad p \in \Pi_i^{2,SO}, i \in I. \quad (12b)$$

For a route  $p$ , with  $f_p^{SO} > 0$ ,  $\rho_p$  and  $\sigma_p$  will describe the amount, negative and positive respectively, by which the current route toll, given by  $\sum_{a \in A} \tau_a \delta_p^a$ , deviates from the first-best route toll. For a route  $p$  with  $f_p^{SO} = 0$ , the negative deviation from the first-best route toll is instead only given by the maximum deviation in OD-pair  $i$ ,  $\mu_i$ . This will make it possible to later iteratively generate the constraints for unused routes, without introducing additional variables. If  $\rho$ ,  $\sigma$  and  $\mu$  are zero, the corresponding toll vector  $\tau$ , which satisfies (12a) and (12b), will be a valid first-best toll vector. By penalizing the deviation from first-best pricing, i.e. the values on  $\rho_p$ ,  $\sigma_p$  and  $\mu_i$ , a minimization problem can be formulated, with the optimal solution equal to the toll locations and levels which minimize the penalized deviation from first-best route tolls. The problem of minimizing the penalized deviation (PD) is formulated as

$$\min_{\rho, \sigma, \mu, \tau, y} z = \beta_1 \sum_{i \in I} \sum_{p \in \Pi_i^{1,SO}} f_p^{SO} (\rho_p + \sigma_p) + \beta_2 \sum_{i \in I} q_i^{SO} \mu_i \quad (13a)$$

$$\sum_{a \in A} (\alpha t_a(v_a^{SO}) + \tau_a) \delta_p^a + \rho_p - \sigma_p = \pi_i^{SO}, \quad p \in \Pi_i^{1,SO}, i \in I \quad (13b)$$

$$\sum_{a \in A} (\alpha t_a(v_a^{SO}) + \tau_a) \delta_p^a + \mu_i \geq \pi_i^{SO}, \quad p \in \Pi_i^{2,SO}, i \in I \quad (13c)$$

$$\rho_p \leq \mu_i, \quad p \in \Pi_i^{1,SO}, i \in I \quad (13d)$$

$$(\tau, y) \in T^V, \quad (13e)$$

$$\rho_p \geq 0, \sigma_p \geq 0, \quad p \in \Pi_i^{1,SO}, i \in I \quad (13f)$$

$$\mu_i \geq 0, \quad i \in I. \quad (13g)$$

Constraint (13d) is introduced to make it possible to develop an efficient solution algorithm, and the practical interpretation of (13d) is that the negative deviation from first-best route tolls for a route with positive flow cannot exceed the maximum negative deviation for the routes with zero flow in the same OD-pair. This is further discussed in the next section.

In the objective function (13a), for any route with positive flow,  $p \in \Pi_i^{1,SO}$ , the deviation from first-best route tolls is weighted by a constant  $\beta_1$  and the SO route flow  $f_p^{SO}$ . For each OD-pair,  $i \in I$ , the maximum deviation from first-best route tolls is weighted by the constant  $\beta_2$  and the SO travel demand  $q_i^{SO}$ . The weights reflect that, for routes with positive flows, it is reasonable that a small deviation from first-best route tolls is more important for routes with large flows, compared with routes with smaller flows. For routes with zero flow in the SO solution, the same

argument is used when weighting  $\mu_i$  by  $q_i^{SO}$ , to reflect the importance of OD-pairs with a high SO demand. If the first-best tolls can be completely recreated on a route level, the resulting toll vector is a first-best solution, with the optimal objective function value  $z^*=0$ .

For fixed toll locations an equivalent minimization problem can be formulated as

$$\min_{\rho, \sigma, \mu, \tau} z = \beta_1 \sum_{i \in I} \sum_{p \in \Pi_i^{SO}} f_p^{SO} (\rho_p + \sigma_p) + \beta_2 \sum_{i \in I} q_i \mu_i \quad (14)$$

subject to  $\tau \in T^F$  and constraints (13b)-(13g)

Note that the link flows are assumed to be fixed (to SO flows), and the optimization problem either takes the form of an LP (if the toll locations are fixed) or a MILP (if the toll locations are variable). For optimal solutions to (13) with optimal objective function values close to zero this is reasonable and the toll level solution will be close to the global maximizer of (9). As the number of tollable links is reduced, the difference between the true equilibrium link flows and the system optimal ones will increase, and the toll level solution to (13) can be expected to move further away from the global maximizer of (9). The benefit, on the other hand, is the possibility to find good toll location and toll levels for large networks in reasonable time.

Even if the set of used routes is known, it is still a matter of expressing (11b) for every unused route, which can be expected to be large. Therefore, a cutting constraint algorithm (CCA) is used, both for solving the LP and the MILP. For the MILP case, the problem, however, becomes too large to be solved to optimality for larger networks, and a greedy heuristic is adopted for the example of the Stockholm network presented in Section 5.

While (11) is formulated based on route flows, a corresponding link based model can be formulated (Yildirim and Hearn, 2005). Since the link based formulation will not rely on explicitly formulating the set of used and used routes, it can directly be implemented and solved with any commercially available solver. The major limitation with adopting a link based formulation is the number of constraints introduced. A link based formulation will approximately have the number of constraints equal to the number of links multiplied with the number of OD-pairs, while the route based version, adopted in this paper, will have the number of constraints equal to the number of used routes, plus the number of additionally added constraints for unused routes. Generating the unused routes iteratively with a CCA, it is thus possible to keep the number of constraints considerably smaller compared with the link based formulation.

One potential problem is the non-uniqueness of route flows. In contrast to the SO link flow solution, the SO route flow solution is not unique, either in terms of used routes or flows on the routes. For the case of (11), the set of first-best toll vectors will not depend on the SO route flow solution. When formulating (13), the objective function will, however, depend on the SO route flow solution (through the weighting parameters). Thus, the SO route flow solution can affect the resulting objective function value, as well as the computed toll levels and locations. While this is clearly a potential limitation of the presented approach, it has for the example of the Stockholm network (used in Section 5 for numerical results) been shown that in practice the choice of SO route flow solutions has a negligible effect on the results.

#### 4. Solution approach for large networks

Constraint (13b) is formulated for every route with a positive flow in the SO solution, and constraint (13c) is formulated for every route in the network with zero flow in the SO solution. The total number of routes with zero flow will for real world traffic networks be large, and to generate them all a priori is not practical possible. Also, it not expected that every constraint in



(13c) will be binding in the optimal solution. Thus, constraints in (13c) can be generated iteratively when solving both (13) and (14).

#### 4.1 Fixed toll locations

Let us first consider the case with fixed toll locations. To solve (14), a reduced set of constraints for the routes with zero flow is formulated,  $\bar{\Pi}_i^{2,SO}$ , which is a (possibly empty) subset of  $\Pi_i^{2,SO}$ .

(14) can then be formulated as

$$\min_{\rho, \sigma, \mu, \tau, y} z = \beta_1 \sum_{i \in I} \sum_{p \in \Pi_i^{1,SO}} f_p^{SO} (\rho_p + \sigma_p) + \beta_2 \sum_{i \in I} q_i^{SO} \mu_i \quad (15a)$$

$$\sum_{a \in A} (\alpha t_a (v_a^{SO}) + \tau_a) \delta_p^a + \rho_p - \sigma_p = \pi_i^{SO}, \quad p \in \Pi_i^{1,SO}, i \in I \quad (15b)$$

$$\sum_{a \in A} (\alpha t_a (v_a^{SO}) + \tau_a) \delta_p^a + \mu_i \geq \pi_i^{SO}, \quad p \in \bar{\Pi}_i^{2,SO}, i \in I \quad (15c)$$

$$\rho_p \leq \mu_i, \quad p \in \Pi_i^{1,SO}, i \in I \quad (15d)$$

$$(\tau, y) \in T^V, \quad (15e)$$

$$\rho_p \geq 0, \sigma_p \geq 0, \quad p \in \Pi_i^{1,SO}, i \in I \quad (15f)$$

$$\mu_i \geq 0, \quad i \in I. \quad (15g)$$

Let  $(\rho^*, \sigma^*, \mu^*, \tau^*)$  be the optimal solution to (15). Then, the search for a violated constraint in (14), for OD-pair  $i$ , can be formulated as

$$\min_{p \in \Pi_i^{2,SO}} w_i = \sum_{a \in A} (\alpha t_a (v_a^{SO}) + \tau_a^*) \delta_p^a + \mu_i^*. \quad (16)$$

To only search for routes in  $\Pi_i^{2,SO}$ , which minimize (16), is not trivial since it will require a complete enumeration of the routes with zero flow in each OD-pair. Since constraint (13d) is included in (14), any route in  $\Pi_i^{1,SO}$ , will also satisfy  $\sum_{a \in A} (\alpha t_a (v_a^{SO}) + \tau_a) \delta_p^a + \mu_i \geq \pi_i^{SO}$ . Thus, (16) can be solved by finding the shortest path in each OD-pair (since  $\mu_i^*$  is constant), with link costs given by  $\alpha t_a (v_a^{SO}) + \tau_a^*$ . Let  $w_i^*$  be the optimal objective function value to (16). If  $w_i^* < \pi_i^{SO}$ , then there exists a route with zero flow which makes the current solution infeasible if it would be included in  $\bar{\Pi}_i^{2,SO}$ .

The CCA for solving (15) can now be formulated as:

- Step 0. For each OD-pair  $i \in I$ , initiate the set of routes with zero flow  $\bar{\Pi}_i^{2,SO}$ .
- Step 1. Solve (15), with optimal solution  $(\tilde{\tau}, \tilde{\mu}, \tilde{\rho}, \tilde{\sigma})$  and objective function value  $\tilde{z}$ .
- Step 2. Solve (16) for each OD-pair  $i \in I$ , with optimal objective function value  $w_i^*$  and corresponding optimal solution  $p^*$ . If  $w_i^* < \pi_i^{SO}$ , set  $\bar{\Pi}_i^{2,SO} := \bar{\Pi}_i^{2,SO} \cup p^*$ . If  $w_i^* \geq \pi_i^{SO}$ , for every OD-pair  $i \in I$ , terminate the algorithm, otherwise continue with Step 1.

In each iteration of the CCA, at least one route is added, or it is concluded that the solution to (15) also solves (14). Since every link has a positive cost, there will be no routes which include cycles, and thus, the algorithm must terminate in a finite number of iterations.

#### 4.2 Variable toll locations

For variable toll locations the optimization problem (13) takes the form of a MILP. Using the CCA presented in the previous section would require one MILP to be solved in each iteration, which is not practical for larger networks. Also, for the numerical results presented for the Stockholm network in the next section, it has not been possible to solve the MILP to optimality even when  $\beta_2$  is set to zero, and no unused routes need to be generated. Thus, to show the applicability of using (13) to find toll locations and levels which provide good solutions to (9), a greedy heuristic has been developed. The purpose of the greedy heuristic is to provide good, but not necessarily optimal, solutions to (13), in order to evaluate the approach of minimizing the deviation from first-best route tolls.

Consider (8), and let  $\lambda_a$  be the dual variable corresponding with the constraint  $\tau_a \leq \tau_a^U g_a$  in (7). The value of the dual variable gives an estimate on how much the objective function value would improve by a unit change of the right hand side. Thus, for a link  $a$  with  $g_a = 0$ ,  $\lambda_a \tau_a$  will give an estimate on the potential improvement of the objective function value from introducing a toll equal to  $\tau_a$  on link  $a$ . While  $\lambda_a$  is not likely to be valid for the whole range from 0 to  $\tau_a^U$ , and it is not known what the actual value on  $\tau_a$  would be if link  $a$  is actually tolled,  $\lambda_a$  can still be used as an estimate on the importance of tolling link  $a$ . In the greedy heuristic, links are chosen iteratively to be included in the solution based on the  $\lambda$ -values, and after a link is added (14) is resolved in order to update the  $\lambda$ -values based on the currently selected toll locations.

The greedy algorithm can more formally be written as

- Step 0. *Initiate* by setting  $g_a := 0$  for every link  $a \in A$
- Step 1. *Solve TLP* in (14) to obtain  $\lambda$
- Step 2. *Add toll location* Find toll location  $b = \underset{a}{\operatorname{argmin}} (\lambda_a (1 - g_a))$ , and set  $g_b := 1$ .
- Step 3. *Terminate* algorithm if  $\sum_{a \in A} g_a = k$ , otherwise continues with Step 1.

A low toll level may indicate less important toll locations, and to further improve the solution quality, the greedy algorithm is rerun with the toll locations with a toll level  $\tau_a$  below the threshold  $\kappa$  removed. This provides a mechanism to remove toll locations which seemed important in the early iterations, but which in the end turned out to give a small contribution to reducing the objective function value. Removing toll locations with  $\tau_a < \kappa$  can be repeated several times with  $\kappa$  reduced by a parameter  $\psi$  each time.

#### 4.3 Bounding the minimum toll booth solution

For comparison, it is of interest to obtain a good estimate on the minimum number of toll locations required to reach SO link flows and demands. Dropping (11b) from (11) results in the relaxed set of first-best toll vectors

$$\sum_{a \in A} (\alpha t_a^{SO} + \tau_a) \delta_p^a = \pi_i^{SO}, \quad p \in \Pi_i^{1,SO}, i \in I. \quad (17)$$

A relaxation of the minimum toll booth problem (Hearn and Ramana, 1998) can then be formulated as

$$\begin{aligned} & \min_y \sum_{a \in A} y_a \\ & \text{subject to } \sum_{a \in A} (\alpha t_a^{SO} + \tau_a) \delta_p^a = \pi_i^{SO}, \quad i \in I, p \in \Pi_i^{1,SO}. \end{aligned} \quad (18)$$

The solution to (18) will give an underestimation of the number of tolls required to achieve first-best pricing. Using the greedy heuristic, but continuing to add toll locations until the optimal objective function value of (14) is below some threshold close to zero<sup>2</sup>, will result in an upper bound on the number of tolls required for achieving first-best pricing.

## 5. Numerical results

The approach for finding good toll locations and toll levels in a congestion pricing scheme, based on the PD approach, has been applied to network models of Sioux Falls and Stockholm. The Sioux Falls model is well used in research papers addressing optimal network design and optimal pricing schemes. The version of the Sioux Falls network adopted in this paper is the elastic demand model, first presented in Yildirim (2001), and later used in Ekström et al. (2013) for evaluating a global optimality approach when optimizing toll locations and toll levels. The Stockholm network has previously been used in Ekström et al. (2014), to study optimal toll locations and corresponding toll levels with a heuristic approach based on a smoothening technique.

To solve both (13) and (14) require a number of LPs to be solved (problems (15) and (16)), and for the numerical results presented here, the commercially available solver CPLEX version 12.2 (IBM, 2010) has been used.

### 5.1 The Sioux Falls network

The version of the Sioux Falls network model, used in this paper, has 79 links and 30 OD-pairs. The link travel time functions are on the BPR-form (Bureau of Public Roads, 1964) and the complete network data is given in Yildirim (2001). For the Sioux Falls network all costs and tolls are given in the unit of minutes, and applying MSCP tolls results in an improvement of the social surplus by 2,722 minutes. In Ekström et al. (2013), a global optimization approach is applied to find the number of optimal toll locations, and their location and corresponding optimal toll levels, given a cost for locating each toll. While the number of toll locations are variable in Ekström et al. (2013), the number of tolls to locate is fixed in (9). To be able to compare the results, the number of tolls to locate will therefore be given by the resulting number of toll locations from Ekström et al. (2013). For the Sioux Falls network model, (13) can be solved to optimality by applying the CCA directly, i.e. by letting the toll locations be variable in (15). This is only possible for small network model, for which the resulting MILP is affordable to solve to optimality in each iteration of the CCA.

For all experiments  $\beta_1$  is set equal to one and  $\beta_2$  is varied between 0 and 1.5 in steps of 0.5. In Table (1), the resulting improvement in social surplus is presented for 7, 11 and 14 number of located tolls ( $k$ ), and the results are compared with the best found solution from using the global optimization approach in Ekström et al. (2013). From the results it is clear that for this limited analysis, the approach presented in this paper perform well in comparison to the results from using the global optimization approach. The global optimization approach should in theory be able to find the global optimal solution. In practice, the computational time required to prove optimality of a solution is large, and therefore the global optimization approach is applied with a time limit. Thus, global optimality is in practise usually not proven for the solutions from this approach. For 11 located tolls, this is clearly a case when the global optimization approach perform worse, compared with the approach presented in this paper. Also, except for  $\beta_2=0$  the solution is robust in terms of how the weighting parameters are set.

---

<sup>2</sup> For larger networks there will be some convergence error when solving the user equilibrium problem and thus it is not appropriate to use zero as termination criteria.

**Table 1. Resulting improvement in social surplus for the Sioux Falls network.**

k	$\beta_2=0.0$	$\beta_2=0.5$	$\beta_2=1.0$	$\beta_2=1.5$	Global optimality approach
7	- 141	2,281	2,316	2,259	2,506
11	1,823	2,573	2,573	2,562	2,531
14	1,835	2,672	2,672	2,669	2,698

In order to evaluate the greedy algorithm it has been applied with  $\beta_1=\beta_2=1$ . In Figure 1, a comparison of the improvement in social surplus is given for both the 7, 11 and 14 located tolls presented in Table 1, and additionally for 5, 10, 15, 20 and 25 located tolls. First of all the results presented in Figure 1 show that the benefit from actually solving (13) to optimality is large, compared with using the greedy algorithm. It is also clear that the differences between the two solution methods diminish when the number of located tolls increases. It is especially interesting to see that the SO solution is reached with 26 located tolls, and that the greedy algorithm requires 27 located tolls in order to reach the SO solution. In terms of computing the minimum toll booth solution, the greedy heuristic performs well. These results can also be compared with the minimum toll booth solution from Yildirim (2001) which provide a minimum toll booth solution with 28 located tolls, based on solving the (MILP) link based minimum toll booth problem. In Yildirim (2001) the solution algorithm is terminated before an optimal integer solution has been verified due to excessive computational time, although a solution with 26 located tolls is found by ad-hoc means. Formulating the set of feasible first-best toll vectors based on route flows rather than link flows, and solving the minimum toll booth problem based on the same CCA as described here, results in the optimal solution of 26 located tolls within seconds, which suggests that in terms of computational efficiency the route based version is superior.

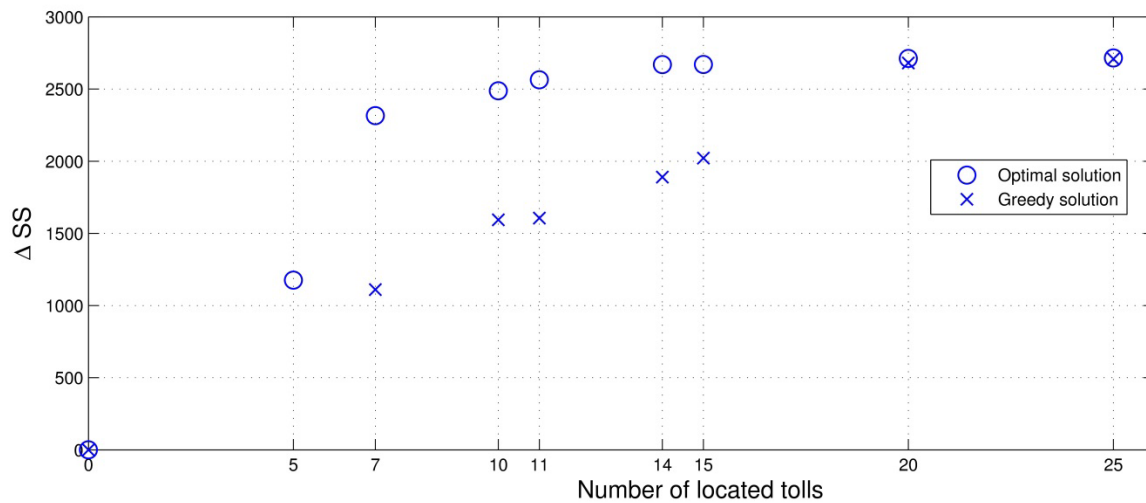


Figure 1. Comparison between optimal and greedy solution of (13), based on the improvement in social surplus.

### 5.2 The Stockholm network model

The Stockholm network used in this paper has 392 links (312 if the connectors to origin and destination zones are excluded) and 40 zones, resulting in 1,560 OD-pairs. Some pairs of links are only used to give a realistic graphical representation of the network, and can be replaced with one single link, which reduce the number of tollable links required for MSCP to 291 links. The link travel time functions are given on the form

$$t_a(v_a) = \frac{v_a}{p_{1,a}n_a} + p_{2,a} \left( 1 + \left( \frac{v_a}{p_{3,a}n_a} \right)^{p_{4,a}} \right) L_a,$$

where  $L_a$  is the link length,  $n_a$  the number of lanes, and  $p_{1,a} - p_{4,a}$  link type specific parameters.

The practical road capacity is given by  $p_{3,a}n_a$ .

The demand model describes the choice between public transport and car for the travellers with access to car during the morning rush hour. The choice between public transport and car is given by a binomial logit model. Assuming that the travel cost for public transport is not depending on the number of travellers using it, the pivot point version of the binomial logit model Kumar (1980) can be used, which for OD-pair  $i$  takes the form

$$q_i = D_i(\pi_i) = T_i \frac{A_i}{A_i + K_i e^{\eta(\pi_i - \pi_i^0)}}, \quad (19)$$

Where  $T_i$  is the total travel demand in OD-pair  $i$ , and  $A_i$  and  $K_i$  are the demand for car and public transport in the non-tolled scenario. The minimum travel cost in OD-pair  $i$ , in the non-tolled scenario, is given by  $\pi_i^0$  and  $\eta > 0$  is the dispersion parameter. The number of travellers by car in the non-tolled scenario is always less than  $T_i$  and when  $\pi_i$  is increased  $q_i$  will decrease.

The Stockholm network represents an aggregated traffic network of the Stockholm region (Figure 2). The demand model (19) is based on data from the demand forecast model T/RIM (Engelson and Svalgård, 1995). The T/RIM model is, however, calibrated for a full Stockholm network, with about 1,100 links and 1,250 zones. In this paper an aggregated version of the T/RIM model is used, and it is possible that using the aggregated traffic network together with T/RIM data, without further calibrations, will result in higher link flows compared with results from other models for the Stockholm region. In Transek (2003) several alternative models are compared for Stockholm, and for the non-tolled scenario the flow across the current cordon in Stockholm vary between 38,357 and 47,922 vehicles per hour, while the aggregated model used in this paper results in 41,731 vehicles per hour. For the purpose of evaluating the PD approach presented in this paper, the aggregated Stockholm network is considered to be a good example of a real network model. The car occupancy  $\chi = 1.13$  travelers per car and the dispersion parameter  $\eta = 0.07$ , are also provided from the T/RIM model, and the value of time is set to 1.2 SEK per minute.

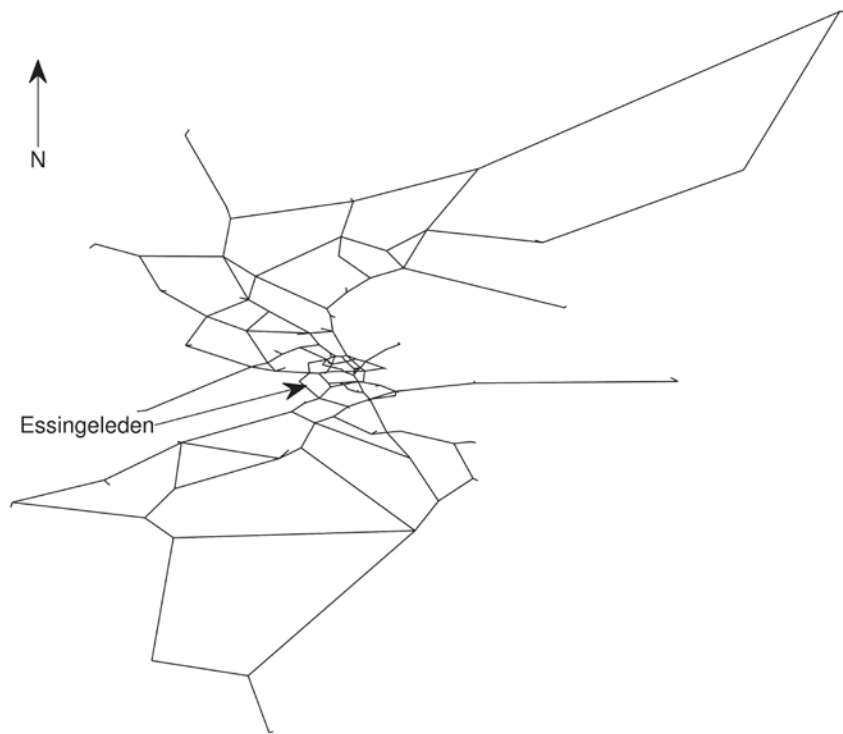


Figure 2. The Stockholm network.

The currently implemented congestion pricing scheme in Stockholm is presented in Figure (3), for a cut-out of the central parts of Stockholm. In the aggregated network there are a total of 20 toll facilities located for the currently implemented scheme, to be compared with the 37 actual located toll facilities. For each located toll, 20 SEK is charged each car passing the toll facility during peak hour. For the SO link flow and demand distribution, which can be achieved by the MSCP tolls,  $\tau^{\text{MSCP}}$ , the improvement in social surplus is 1,031,499 SEK per rush hour, which can be compared with the improvement in social surplus achieved by the currently implemented cordon which is 303,715 SEK per rush hour.

When the demand is given by (19), the user equilibrium problem with elastic demand (2) can be solved by using the partial linearization algorithm presented in Evans (1976), in which a series of fixed demand user equilibrium problems are solved iteratively. In this paper the partial linearization algorithm is used together with the Disaggregated Simplicial Decomposition (DSD) algorithm (Larsson and Patriksson, 1992) for solving each fixed demand problem. The benefit from using the DSD algorithm is the availability of route information, which is needed when formulating (13). Also the DSD algorithm has a re-optimization capability which is a useful feature when solving a series of similar fixed demand user equilibrium problems. Any algorithm providing explicit route information can, however, be used.

For a large traffic network, the convergence when solving the user equilibrium problem will not be perfect, and there may exist routes with positive flow but with the route travel cost differing from the minimum OD-travel cost. In this paper this is handled by removing any route with  $f_p < 0.1$  from the set of used routes.

While the optimization problem solved in Ekström et al. (2014) includes the cost of locating the toll collection facilities, and the number of toll locations is variable in the problem, the resulting toll locations from Ekström et al. (2014) can still be used as comparison to the ones computed with the PD approach presented in this paper.

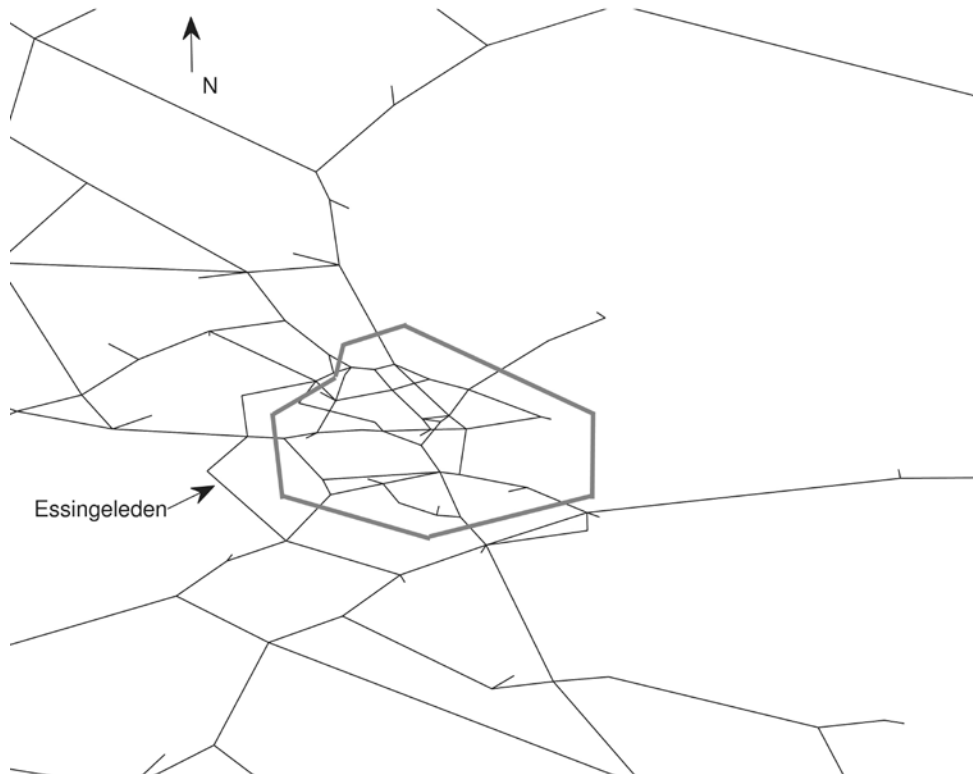


Figure 3. The current congestion pricing cordon in Stockholm.

#### Fixed toll locations

For the experiment presented in this section,  $\beta_1$  is set equal to 1 for all experiments, and  $\beta_2$  is varied between 0 and 2. For  $\beta_2=0$ , no iterations are needed with the CCA since routes with zero flow will not affect the optimal solution, for  $\beta_2>0$ , the CCA generate between 542 and 943 additional routes, during 2 to 4 iterations. The solution time is between 2-20 CPU-seconds, using one Intel P8600 2.4GHz processor.

Using the sensitivity analysis based ascent method, presented in Ekström et al. (2009), optimal toll levels were computed in Ekström et al. (2014) for the currently implemented cordon in Stockholm, as well as for an extended Stockholm cordon, in which the bypass highway "Essingeleden" is also tolled (resulting in a total of 22 tolled links). The toll level solution obtained from the ascent method is denoted  $\tau^{AS}$  with corresponding change in social surplus  $\Delta SS(\tau^{AS})$ , and for comparison these results are presented in Table 2.

The toll level solution which solves (14) is denoted  $\tau^{PD}$ , and applied as toll level solution in (9) results in an improvement of the social surplus by  $\Delta SS(\tau^{PD})$ . The improvement in social surplus associated with the toll levels obtained PD approach are presented in Table 3 and 4 for the current and extended cordon respectively. For comparison,  $\Delta SS(\tau^{PD}) / \Delta SS(\tau^{AS})$  is also given in these tables.

**Table 2. Results (in SEK) for the Stockholm network, from Ekström et al. (2014).**

Scheme	$\Delta SS$	$\tau^{AS}$
Current cordon	303,715	20
Current cordon optimized	399,837	8.1-42.8
Extd. current cordon	430,503	20
Extd. current cordon optimized	583,803	8.7-47.9

**Table 3. Results obtained when solving (14) for the current Stockholm cordon.**

$\beta_2$	$\Delta SS(\tau^{PD})$ (in SEK)	$\frac{\Delta SS(\tau^{PD})}{\Delta SS(\tau^{AS})}$
0	357,000	0.89
0.5	383,181	0.96
1	390,664	0.98
1.5	393,913	0.99
2	392,769	0.98

**Table 4. Results obtained when solving (14) for the extended Stockholm cordon.**

$\beta_2$	$\Delta SS(\tau^{PD})$ (in SEK)	$\frac{\Delta SS(\tau^{PD})}{\Delta SS(\tau^{AS})}$
0	379,631	0.65
0.5	550,466	0.94
1	566,664	0.97
1.5	578,541	0.99
2	574,813	0.98

For  $\beta_2=1.5$ , the toll levels obtained by minimizing the deviation from first-best route tolls reaches 99% of  $\Delta SS(\tau^{AS})$ . While the ascent method requires several hours in computational time, the time required to solve (14) is between 2-20 seconds depending on the number of iterations with the CCA. For all evaluated choices of  $\beta_2$ , it is only  $\beta_2=0$  which performs poorly, and for all other choices the results are close to what is achieved by the ascent method. First of all, this suggests that minimizing the deviation from first-best route tolls may result in toll levels close to a local optimal solution, and secondly, that the PD approach is a practical useful approach for minimizing (9). The results also suggest that the PD approach is robust in terms of values on the  $\beta$ -parameters, and the relative small number of generated routes and the low computational time suggests that the approach will be applicable for even larger network models.

To evaluate how the choice of route flow solution used for defining the set of feasible first-best toll vectors can affect the performance of the PD approach, three additional route flow solutions, with different properties, have been used when applying the PD approach to the current Stockholm cordon. A linear program can be formulated with the feasible region defining the route flows which realise the SO link flow and demand solution. Different objectives can then be applied in order to compute route flow solutions with different properties. The complete set of unused routes has not been included in this analysis, but a subset of the unused routes (extracted from the DSD algorithm) has been included. Route set 1 is the initial set of route from the DSD algorithm. Set 2 is obtained by minimizing the sum of the route flows on the ten routes with largest flow in the initial solution. By maximizing the sum of the route flows on the unused route (with travel cost equal to the minimum OD travel cost) in the initial solution, route set 3 and 4 are obtained, with the difference that in route set 4 no used route in the initial solution is allowed to be reduced to less than 50% of the flow in the initial solution.

The four different route flow solutions are presented in Table 5, together with the optimal objective function value ( $z^*$ ) of (14) and resulting improvement in social surplus ( $\Delta SS$ ) for the example of  $\beta_1 = \beta_2 = 1$ . Let  $P$  be the set of considered equilibrium routes (for both used and unused routes), and the initial and new route flow solution denoted  $f^{init}$  and  $f^{new}$  respectively. The distance between the initial and new route flow solution is then expressed as  $D = \sum_{p \in P} |f^{new} - f^{init}|$ , and is also presented in Table 5.



**Table 5. Comparison of results from using alternative route flow solutions for defining the set of first-best toll vectors**

Route set	Number of routes with positive flow	$z^*$	$\Delta SS$ (in SEK)	$D$
1	2,527	2,638,941	390,664	0
2	1,601	2,638,040	391,001	32,179
3	1,618	2,634,140	388,182	32,116
4	2,707	2,637,940	390,161	16,601

From Table 5 it is clear that for this example the choice of route flow solution has negligible effect on the toll level solution. The conclusion is the same for other choices of values on  $\beta_1$  and  $\beta_2$ , and for the extended Stockholm cordon as well. These results are therefore excluded from this presentation.

#### *Variable toll locations*

The greedy heuristic presented in Section 4.2 has been applied to solve (13). While the greedy heuristic is not expected to solve (13) to optimality, it provides solutions which can be used for evaluating the PD approach of minimizing the deviation from first-best route tolls in order to obtain good toll locations and toll levels to (9). Results are presented for choosing 24, 43, 69, 120, 160 and 200 tollable links out of 291 possible toll locations. The  $\kappa$ -parameter in the greedy heuristic is scenario specific and set to 20, 10, 5, 3, 2 and 1 respectively, and the reduction factor has been set to  $\psi = \kappa/10$ , resulting in a total of 10 reruns with the greedy heuristic for each scenario.

The first three scenarios (24, 43 and 69 tollable links) correspond with solutions from using the smoothening heuristic presented in Ekström et al. (2009). While the smoothening heuristic is used for a variable number of toll locations, in order to maximize the social surplus minus the cost of setting-up and operating the toll collection facilities, the resulting toll locations will also provide good solutions for the case when the number of toll locations are fixed to the optimal number of located tolls from the smoothening heuristic. In this paper, each of these scenarios will be used as comparison, when searching for the optimal toll locations given the number of tolls to locate from the smoothening heuristic. The scenario with 69 located tolls is obtained from Ekström et al. (2014), in which the set-up and operational cost is estimated to 500 SEK per lane of each link. To obtain 24 and 43 located tolls, the set-up and operational cost is set to 10,000 SEK and 5,000 SEK respectively for each link. Note that for the purpose of this paper, the cost associated with a toll location is not relevant, and is just set to a value which results in an appropriate number of tolls being located. For 120, 160 and 200 located tolls there exist no comparison from the smoothening heuristic and results are presented for these scenarios to show the performance of the PD approach when the number of tolled links is increased. Lower and upper bound estimations on the number of toll locations required for first-best pricing is obtained by the approach presented in Section 4.3, which results in a lower bound of 211 tolls<sup>3</sup> and an upper bound of 219 tolls<sup>4</sup>. Computing the improvement in social surplus with the 219 toll locations results in the same improvement of the social surplus as is reached with MSCP tolls.

In Figure 4, the final objective function value of (13), from the greedy heuristic, is given as a function of the number of located tolls for each scenario. Let  $\tau^{\text{PD}}$  and  $\tau^{\text{SH}}$  denote toll level solutions from the PD approach and the smoothening heuristic respectively. In Table 6  $\Delta SS(\tau^{\text{PD}})$  is presented together with  $\Delta SS(\tau^{\text{PD}})/\Delta SS(\tau^{\text{SH}})$  and  $\Delta SS(\tau^{\text{PD}})/\Delta SS(\tau^{\text{MSCP}})$  for comparison. As an alternative approach for determining good toll locations, the  $k$  number of

<sup>3</sup> The solution to (18) is provided by running CPLEX version 12.2 for 8 hours

<sup>4</sup> The upper bound is obtained by adding toll locations until  $\hat{z} = 17.63$ , which is the value on  $\hat{z}$  when every link is tollable.

links with the largest MSCP toll levels are chosen to be tolled, with toll level equal to the MSCP link toll. The results from this approach is denoted as the k-MSCP solution, and the corresponding change in social surplus ( $\Delta SS(\tau^{k\text{-MSCP}})$ ) is given in Table 6 for comparison.

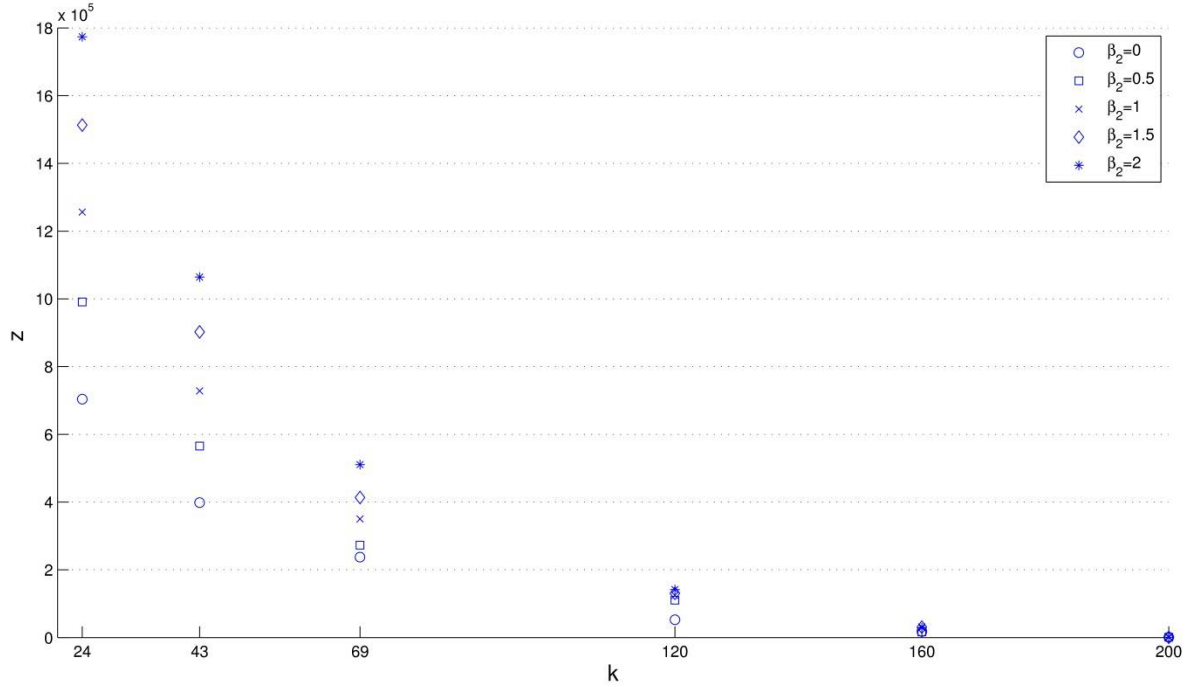


Figure 4. Objective function value when solving (13) with the greedy heuristic.

From Figure 4 it is clear that  $\hat{z}$  is reduced when the number of located tolls is increased. In Table 6, it can be seen that the highest value on  $\Delta SS(\tau^{PD})$  is obtained with  $\beta_2=1$  for  $k=24$ , with  $\beta_2=1.5$  for  $k=43$ , and with  $\beta_2=0.5$  for the remaining scenarios. For  $k=160$  and  $k=200$ , the choice of  $\beta_2$  seems to be less important, and  $\beta_2=0$  in general performs poorly. It is, however, difficult to make any definitive statement based on the solutions from the greedy heuristic, since optimality is not guaranteed. Comparing the best toll locations for each scenario, based on  $\Delta SS(\tau^{PD})$ , with the solution obtained by the smoothening heuristic, it is clear that  $\Delta SS(\tau^{PD}) < \Delta SS(\tau^{SH})$  for all scenarios. It should, however, be noted that the computational time for the PD approach is between 100 and 7,960 seconds, depending on the number of located tolls, while the computational time required by the smoothening heuristic to provide the solutions is counted in days. For the cases with 24, 43 and 69 tollable links, the maximum computational time required by the PD approach is 1,156 seconds, and the resulting improvement of the social surplus is within the range 87% to 94% of what is achieved with the smoothening heuristic. Comparing the PD approach with the k-MSCP approach shows that when the number of located tolls is increased, the difference between the two approaches diminishes. For up to 69 located tolls, the PD approach, however, clearly outperforms the k-MSCP approach.

For the best found solution for each number of located tolls, the ascent method is applied to further polish the solution. The resulting toll levels are denoted  $\tau^{PD-AS}$  with the change in social surplus given by  $\Delta SS(\tau^{PD-AS})$ , and these results are presented in Table 7 together with  $\Delta SS(\tau^{PD-AS}) / \Delta SS(\tau^{SH})$  and  $\Delta SS(\tau^{PD-AS}) / \Delta SS(\tau^{MSCP})$  for comparison.

The results in Table 7 show that with 24 located tolls (approximately 11% of the number of tolls required to achieve first-best pricing) it is possible to reach 75% of  $\Delta SS(\tau^{\text{MSCP}})$ , and the best obtained toll location solution for 24 located tolls is presented in Figure 5. Allowing 120 toll locations makes it possible to reach 96% of  $\Delta SS(\tau^{\text{MSCP}})$ . With 160 and 200 located tolls, the deviation from  $\Delta SS(\tau^{\text{MSCP}})$  is only minor. In the light of these results it is apparent that by optimizing toll locations and toll levels, attractive congestion pricing schemes can be designed with a significantly reduced number of located tolls compared with first-best pricing. The results also show that for a realistic traffic network, several of the toll locations required to achieve first-best pricing will have negligible contribution to the improvement in the social surplus.

**Table 6. Results from solving (13) with the greedy heuristic, and comparison with k-MSCP.**

$k$	$\beta_2$	$\Delta SS(\tau^{\text{PD}})$	$\frac{\Delta SS(\tau^{\text{PD}})}{\Delta SS(\tau^{\text{SH}})}$	$\frac{\Delta SS(\tau^{\text{PD}})}{\Delta SS(\tau^{\text{MSCP}})}$	$\frac{\Delta SS(\tau^{\text{k-MSCP}})}{\Delta SS(\tau^{\text{MSCP}})}$
24	0	517,850	0.61	0.50	0.55
	0.5	706,708	0.84	0.69	
	1.0	756,639	0.90	0.73	
	1.5	739,082	0.88	0.72	
	2.0	728,636	0.86	0.71	
43	0	650,341	0.69	0.63	0.71
	0.5	823,620	0.87	0.80	
	1.0	803,892	0.85	0.78	
	1.5	827,413	0.87	0.80	
	2.0	766,405	0.81	0.74	
69	0	749,366	0.75	0.73	0.82
	0.5	932,128	0.94	0.90	
	1.0	929,281	0.94	0.90	
	1.5	930,552	0.94	0.90	
	2.0	859,512	0.87	0.83	
120	0	929,168	-	0.90	0.96
	0.5	971,283	-	0.94	
	1.0	968,918	-	0.94	
	1.5	960,847	-	0.93	
	2.0	954,562	-	0.93	
160	0	964,312	-	0.93	0.99
	0.5	1,023,991	-	0.99	
	1.0	1,015,478	-	0.98	
	1.5	1,015,637	-	0.98	
	2.0	1,018,705	-	0.99	
200	0	1,023,500	-	0.99	1.00
	0.5	1,031,453	-	1.00	
	1.0	1,028,560	-	1.00	
	1.5	1,030,953	-	1.00	
	2.0	1,030,909	-	1.00	

**Table 7. Results obtained by using the ascent method for polishing  $\tau^{\text{PD}}$ .**

$k$	$\beta_2$	$\Delta SS(\tau^{\text{PD-AS}})$	$\frac{\Delta SS(\tau^{\text{PD-AS}})}{\Delta SS(\tau^{\text{SH}})}$	$\frac{\Delta SS(\tau^{\text{PD-AS}})}{\Delta SS(\tau^{\text{MSCP}})}$
24	1.0	775,681	0.91	0.75
43	1.5	864,410	0.91	0.84
69	0.5	961,251	0.97	0.93
120	0.5	989,821	-	0.96
160	0.5	1,026,230	-	1.00
200	0.5	1,031,459	-	1.00

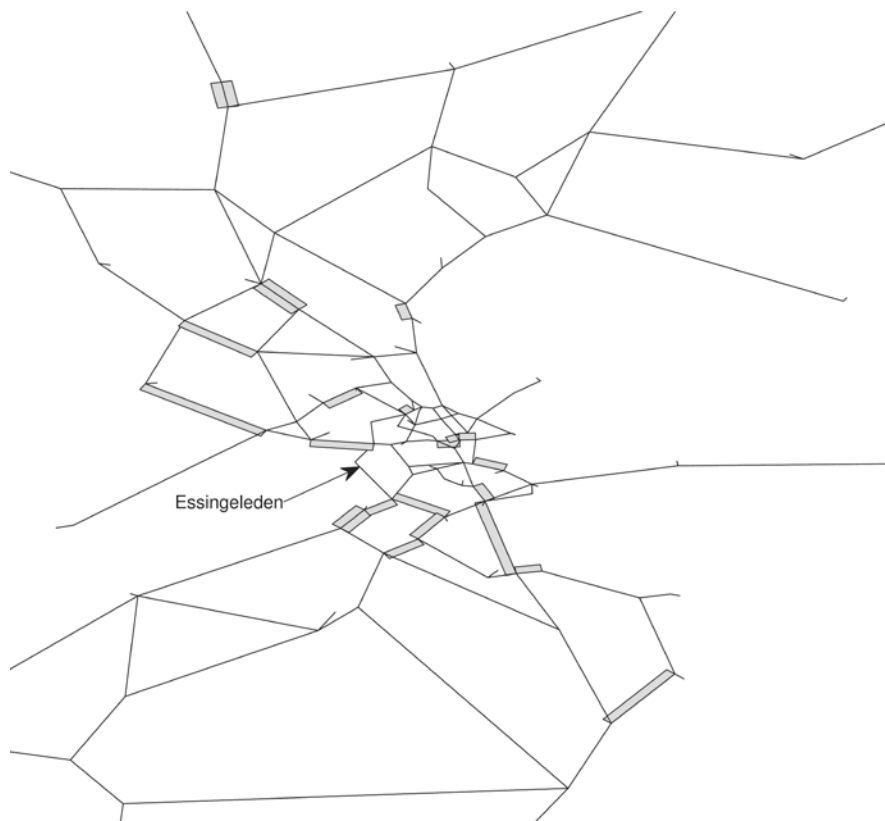


Figure 5. Best obtained toll location solution for 24 located tolls. Tolled links are marked with a grey bar.

## 6. Discussion and further research

In this paper an approach based on minimizing the deviation from first-best route tolls is applied for finding toll locations and levels which maximize the social surplus. In the numerical results it is shown for a realistic traffic network that results can be obtained, for fixed toll locations, with small differences from known local optimal solutions, in very short computational time. For variable toll locations, it is shown that the approach is able to find good solutions, within reasonable computational time.

One limitation of the proposed approach is that the quality of the solution depends on the values of the parameters  $\beta_1$  and  $\beta_2$ . While it may not be possible to know what values to use on these parameters in advance, the numerical results suggests that the solution approach is not very sensitive on the selection of the parameter values. Giving the unused routes a weigh close to zero or double that of the used routes ( $\beta_2 \geq 2\beta_1$ ) clearly show a worse performance of the approach. For all other evaluated  $\beta$ -values, the differences in term of solution quality are small. Another potential problem is the set of used SO routes is not unique, either in terms of route flows or used routes. For the Stockholm network, this has not shown to be an actual problem, and the solution when optimizing the toll levels in the current cordon are not sensitive to the choice of used routes and their flows.

This paper also provides results which give some insight into “close to” first-best pricing schemes. While it can be determined that between 211 and 219 tolls are required to achieve first-best pricing, it is possible to find pricing schemes with 160 located tolls which account for 99% of the improvement in social surplus associated with first-best pricing. Even more interesting are the results for 24 and 43 numbers of tolls to locate, which correspond to 11% and 20%

respectively of the 211 number of tolls which at least are required to achieve first-best pricing in the Stockholm network model. With 24 located tolls it is possible to reach 75% of the increase in social surplus associated with first-best pricing, and for 43 located tolls the corresponding number is 85%. Thus, one can argue that to search for the minimum number of toll locations which achieve first-best pricing is not interesting in practice, when the marginal contribution of as much as 80% of the located tolls is small. Comparing the improvement in social surplus achieved by the current cordon, extended cordon and the 24 optimal located tolls, it is clear that the ability to optimize toll locations can significantly improve the performance of the pricing scheme.

While this paper has shown the applicability of the approach on the rather aggregated Stockholm network, future work need to focus on applications to larger transportation models and on further developing the solution approach for the case of variable toll locations. The comparison between optimal and greedy solutions for Sioux Falls network clearly shows the potential in improving the MILP solution approach.

One interesting extension of the PD approach is to apply it to multiclass traffic networks, in which users are differentiated by their value of time. In a multiclass network, MSCP tolls, in the unit of travel time, is equal for all user groups but the corresponding MSCP money tolls will be differentiated with respect to the value of time in each user group. The PD approach can be used in order to minimize the deviation from the first-best route time toll, to set a common money toll for all groups.

The PD approach only rely on the ability to compute MSCP tolls, and this makes it possible to use the approach together with the trial-and-error algorithm, presented in Yang et al. (2004), for computing MSCP tolls in the absence of demand functions, which allow for usage with networks in which the relationship between demand and travel cost cannot be formulated as a closed form function. In de Palma et al. (2005), approximative MSCP tolls are computed for a dynamic traffic model, and the simplicity of the PD approach makes the approach interesting to extend to dynamic traffic models with time expanded networks.

## Acknowledgements

The presentation of this paper benefited from the comments of two anonymous reviewers.

## References

- Bureau of Public Roads. (1964). *Traffic Assignment Manual*. US Department of Commerce, Urban Planning Division, Washington, DC.
- Chen, M., Bernstein, D. and Spasovic, L. (2004). Toll-design problem with stochastic route choice. *Environment and Planning B: Planning and Design*. 35, 731-742.
- Chen, M. and Bernstein D.H. (2004). Solving the toll design problem with multiple user groups. *Transportation Research Part B*. 38(1), 61-79.
- Connors, R.D., Sumalee, A. and Watling, D.P. (2007). Sensitivity analysis of the variable demand probit stochastic user equilibrium with multiple user-classes. *Transportation Research Part B*. 41(6), 593-615.
- de Palma, A., Kilani, M. and Lindsey, R. (2005). Congestion pricing on a road network: A study using the dynamic equilibrium simulator metropolis. *Transportation Research Part A* 39(7-9), 588-611.
- Ekström, J., Engelson, L. and Rydergren, C. (2009). Heuristic algorithms for a second-best congestion pricing problem. *NETNOMICS: Economic Research and Electronic Networking*. 10(1), 85-102.

Ekström, J., Engelson, L. and Rydergren, C. (2014). Optimal Toll Locations and Levels in Congestion Pricing Schemes: a Case Study of Stockholm. Accepted for publication in *Transportation Planning and Technology*.

Ekström, J., Rydergren, C. and Sumalee, A. (2013) Solving a Mixed Integer Linear Program Approximation of the Toll Design Problem Using Constraint Generation within a Branch and Cut Algorithm. *Transportmetrica A: Transport Science*. Published on-line.

Ekström, J., Sumalee, A. and Lo, H.K. (2012). Optimizing toll locations and levels using a mixed integer linear approximation approach. *Transportation Research Part B* 46(7), 834-854.

Engelson, L. and Svalgård, S. (1995). *The travel demand and land-use model for Stockholm region. Technical report*. Inregia AB, Stockholm.

Evans, S.P. (1976). Derivation and analysis of some models for combining trip distribution and assignment. *Transportation Research* 10(1), 37-57.

Hearn, D.W. and Ramana. M.V. (1998). Solving congestion toll pricing models. In Marcotte, P. and Nguyen, S. (ed), *Equilibrium and Advanced Transportation Modeling*. Kluwer Academic, Norwell.

IBM. (2010). *IBM ILOG CPLEX v12.1 User's Manual for CPLEX*.

Kumar, A. (1980). Pivot point modeling procedures in demand estimation. *Transportation Engineering Journal* 106(6), 647-660.

Larsson, T. and Patriksson, M. (1992). Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Science* 26(1), 4-17.

Larsson, T. and Patriksson, M. (1998). Side constrained traffic equilibrium models-traffic management through link tolls. In Marcotte, P. and Nguyen, S. (ed), *Equilibrium and Advanced Transportation Modeling*. Kluwer Academic, Norwell.

Lawphongpanich, S. and Hearn, D. W. (2004). An MPEC approach to second-best toll pricing. *Mathematical Programming*. 101(1), 33-55.

Migdalas, A. (1995). Bilevel programming in traffic planning: Models, methods and challenge. *Journal of Global Optimization*. 7(4), 381-405.

Pigou, A. C. (1920). *Welth and Welfare*. MacMillan, London.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, New Jersey.

Shepherd, S. and Sumalee, A. (2004). A genetic algorithm based approach to optimal toll level and location problems. *Networks and Spatial Economics*. 4(2), 161-179.

Sumalee, A. (2004). Optimal road user charging cordon design: A heuristic optimization approach. *Computer-Aided Civil and Infrastructure Engineering*. 19(5), 377-392.

Sumalee, A., Connors, R. and Watling, D. (2006). An optimal toll design problem with improved behavioural equilibrium model: The case of the probit model. In Lawphongpanich, S., Hearn, D. and Smith, M. (ed) *Mathematical and Computational Models for Congestion Charging, Volume 101 of Applied Optimization*, Springer, US.

Transek (2003). *Miljöavgifter i Stockholm, analys av nuvarande förslag till utformning. Technical report*, Transek AB, Stockholm.

Tsekeris, T. and Voß, S. (2009). Design and evaluation of road pricing: state-of-the-art and methodological advances. *NETNOMICS: Economic Research and Electronic Networking*. 10(1), 5-52.

Verhoef, E. T. (2002a). Second-best congestion pricing in general static transportation networks with elastic demands. *Regional Science and Urban Economics*. 32(3), 281-310.

Verhoef, E. T. (2002b). Second-best congestion pricing in general networks. Heuristic algorithms for finding second-best optimal toll levels and toll points. *Transportation Research Part B*. 36(8), 707-729.

Yang, H. and Lam W. H. K. (1996). Optimal road tolls under conditions of queueing and congestion. *Transportation Research Part A*. 30(5), 319-332.

Yang, H., Meng, Q. and Lee, D.H. (2004). Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions. *Transportation Research Part B* 38(6), 477- 493.

Yang, H. and Zhang, X. (2003). Optimal toll design in second-best link-based congestion pricing. *Transportation Research Records*. 1857, 85-92.

Yildirim, B. and Hearn, D.W. (2005). A first best toll pricing framework for variable demand traffic assignment problems. *Transportation Research Part B*. 39(8), 659-678.

Yildirim, M.B. (2001). *Congestion Toll Pricing Models and Methods for Variable Demand Networks*, Ph.D. thesis, University of Florida, Gainesville.

Yin, Y. (2000). Genetic-algorithms-based approach for bilevel programming models. *Journal of Transportation Engineering*. 126(2), 115-120.

Yin, Y. and Lawphongpanich, S. (2009). Alternative marginal-cost pricing for road networks. *NETNOMICS: Economic Research and Electronic Networking*. 10(1), 77-83.

Ying, J. Q. and Yang, H. (2005). Sensitivity analysis of stochastic user equilibrium flows in a bi-modal network with application to optimal pricing. *Transportation Research Part B*. 39(9), 769-795.

Zerbe, R. O.J. and Dively, D.D. (1994). *Benefit-cost analysis in theory and practice*. HarperCollins, New York.

Zhang, X. and van Wee, B. (2012). Enhancing transportation network capacity by congestion pricing with simultaneous toll location and toll level optimization. *Engineering Optimization*. 44(4), 477-488.

Zhang, X. and Yang, H. (2004). The optimal cordon-based network congestion pricing problem. *Transportation Research Part B*. 38(6), 517-537.