Network Model for Rural Roadway Tolling with Pavement Deterioration and Repair

Promothes Saha
Department of Civil and Architectural Engineering, University of Wyoming, Laramie, WY

&

Ruoyu Liu, Christopher Melson & Stephen D. Boyles*
Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin, 301 E. Dean Keeton St. Stop C1761, Austin, TX 78712, USA

Abstract: The objective of roadway tolling in rural areas is often tied to revenue generation for roadway maintenance. Thus, rural pricing models should directly incorporate a pavement deterioration and maintenance model. However, the interactions between these models are not simple, because tolls cause traffic diversion, which in turn affects deterioration rates and forecasted revenue. This article describes a rural pricing model which calculates diversion endogenously with a network assignment model. This model captures deterioration rates and pavement condition in the toll-setter's objective function, maximizing long-run net present value of the highway infrastructure. A novel deterioration model is used which is particularly suitable for computational efficiency. The resulting model is discontinuous and nondifferentiable, and involves solving a combinatorial knapsack problem as a subproblem. Thus, a simulated annealing-based algorithm is presented to solve it, in the framework of a new solution method built upon partitioning the feasible region. A demonstration is made using a network representing the state of Wyoming (28 zones, 60 nodes, and 188 links). Sensitivity analyses reveal that although the locations for optimal tolling are relatively stable as demand changes, the revenue collected can vary substantially. Relatively simple models are used throughout for computational reasons, and future research should investigate strategies for incorporating more advanced pavement and network models.

1 INTRODUCTION

In rural areas, many transportation agencies are considering roadway tolls as a mechanism to generate revenue to fund roadway maintenance and repair (M&R). In particular, the growth in global trade in recent decades has sharply increased the share of heavy vehicles on rural freeways—for instance, in the United States, heavy vehicles comprise nearly 60% of traffic volume on certain segments of Interstate 80 (I-80) in Wyoming (Wyoming Department of Transportation, 2008a). The impact of heavy vehicles on pavement deterioration is several orders of magnitude larger than that of passenger vehicles (American Association of State Highway and Transportation Officials, 1993) and a pricing strategy can seek to mitigate these costs. Furthermore, the rural context is different in several other important ways: the relative share of long-distance trips is higher; the roadway network is sparser and there are fewer opportunities to divert onto high-quality facilities; the effect of congestion externalities is much smaller than that of pavement deterioration externalities.

In particular, rural pricing induces a complex set of interactions among several entities: the agency, drivers, and the infrastructure itself. On the one hand, imposing a toll will generate revenue which can be used for pavement maintenance. This specific linkage means that accurately calculating benefits from toll revenue requires explicit connection with an M&R model. At the same time, tolls will impact traffic flows, and capturing these diversion effects is extremely important for public transportation agencies: in addition to the usual travel time and fuel expenditures associated with diversion, the post-toll traffic pattern will determine new...
pavement deterioration rates. If heavy vehicles divert from the tolled facility onto facilities of lesser quality, long-term maintenance expenditures may increase by more than what toll revenue can fund. Thus, to fully evaluate the impacts of a rural pricing policy, one must simultaneously include a network model to represent diversion and rerouting, a pavement M&R model to estimate benefits and costs, and a pricing algorithm to guide the search for effective policies.

As described in Section 2, most existing research has focused on one or two of these models, but the authors believe that all three must be simultaneously considered. Aggregate approaches to measuring maintenance externalities (e.g., “each truck causes X dollars in damage”) are overly simplistic and fail to represent differences in existing pavement quality between and within regions, the maintenance options available to a particular agency, and the effects of diversion within the network onto facilities of varying quality. As Section 5 demonstrates, the appropriate pricing levels depend highly on the demand level, current pavement condition, and other scenario-specific information.

The contributions of this article are (1) an integrated framework including a pricing algorithm, a network assignment model to capture diversion, and a pavement M&R model, and (2) a corresponding solution method based on applying simulated annealing to a partitioning of the feasible region. By representing all of these features simultaneously, the interactions between these components are clearly visible and can be accounted for in toll planning. This framework is demonstrated using data and networks from the rural state of Wyoming. For computational reasons it is impractical to use highly detailed or advanced network and M&R models for these purposes. Because the resulting optimization formulation lacks a favorable structure, identifying approximately-optimal tolls requires many iterations of the network and pavement models to evaluate candidate solutions (hundreds of thousands of times in the numerical results presented in this article). Therefore, the network and maintenance models are necessarily simplistic ones—the contribution of the article lies not with these specific components, but integrating them in a way that provides useful information to toll planners. The authors believe that using highly advanced network assignment algorithms or sophisticated deterioration and maintenance model alone or separately is far less useful than combining simpler versions of both in a unified modeling framework that can immediately capture the interactions between them.

The remainder of this article is organized as follows. Section 2 discusses literature related to the pricing problem and the specific models used. Section 3 then describes each component of the model in detail, along with the integrated framework. The solution algorithm for this formulation is described in Section 4, followed by numerical analysis using the state of Wyoming in Section 5. Finally, Section 6 concludes by summarizing the major contributions and identifying directions for further research.

2 LITERATURE REVIEW

In the academic literature, the problem of “optimal” toll selection generally falls into two categories: profit maximization or welfare maximization. The former is often used to model the aims of private companies profiting from the operation of this infrastructure, whereas the latter is often used by public agencies seeking to price externalities. In general, both objectives lead to bilevel optimization problems, in which the upper level is to maximize profit or welfare subject to a lower level constraint in which travelers minimize their own cost of travel in response (Yang and Bell, 2001). Examples of such models include Labbé et al. (1998), who formulated a bilinear network model for profit maximization when only a subset of links can be tolled; Larsson and Patriksson (1998), who obviated the need for solving a bilevel program by identifying the Lagrange multipliers for a suitable set of side-constraints; Unnikrishnan et al. (2009), who considered the interactions between private, profit-maximizing investors and a public agency wishing to minimize travel time; and Unnikrishnan and Lin (2012), who investigated network design in a bilevel context.

In welfare maximization models, traditionally the focus is on congestion externalities, which were identified as early as Pigou (1920) and Beckmann et al. (1956); see Yang and Huang (2005) or Small and Verhoef (2007) for comprehensive treatments of the subject. In the context of the traditional static traffic assignment problem, if there are no restrictions on tolls, the welfare-maximizing tolls can be easily identified by solving a system-optimal assignment and calculating link marginal prices. Researchers have extended this basic model in a number of directions, including time dynamics (Vickrey, 1969), stochastic network conditions (Boyles et al., 2010a), demand uncertainty (Gardner et al., 2011), driver heterogeneity (Dial, 1999a, b), and restrictions to a subset of links (Sumalee, 2004; Meng and Wang, 2008); this latter extension is known as second-best pricing (Verhoef et al., 1996; Verhoef, 2002).

In rural areas, externalities related to pavement deterioration are more important than those related to congestion (Wyoming Department of Transportation, 2008b). A large variety of pavement deterioration and
maintainance models have been developed. One important distinction is between facility-level models and network-level models. Facility-level models focus solely on a single facility (for instance, a bridge, bridge component, or section of pavement), and seek an optimal maintenance policy, whereas network-level models consider a large system of facilities linked in a way that prevents a simple decomposition by facility (as with a total budget constraint, or statistical dependence in deterioration). Examples of facility-level models can be found in Madanat and Ben-Akiva (1994), Gao and Zhang (2006), Deshpande et al. (2010), Boyles et al. (2010b), and Gao and Zhang (2013) using optimization techniques such as dynamic programming and nonlinear programming. Examples of network-level models include Smilowitz and Madanat (2000), Kuhn (2010), Sathaye and Madanat (2011), and Gao et al. (2012). In principle, network-level models are more realistic, but often computational tractability forces simplifications to be made. Zheng et al. (2012) considered interactions between work zone sequencing and traffic diversion on a network for scheduled maintenance, in a somewhat different context than the long-term planning horizon used for toll analysis. As described in the following sections, the tolling model will require a pavement deterioration and maintenance model to be called thousands of times as a subroutine. As a result, a simple relationship proposed by Mahoney and Jackson (1991) is adapted for this purpose. More sophisticated models, such as those developed by the World Bank (see, for instance, Tsunokawa et al., 2002) would improve the accuracy of this component, but would restrict the number of scenarios that can be explored in the computational budget.

An overview of toll-setting in practice can be found in Holguín-Veras et al. (2006), who performed a comparative analysis of toll facilities in the United States, particularly examining whether tolls could be linked to congestion or pavement deterioration externalities, concluding that although heavy vehicles do pay significantly higher tolls, these tolls “do not seem to be proportional to the road space they consumed, or the pavement deterioration they produce.” These authors considered tolling in both urban and rural setting; this article complements their results by incorporating a network into the rural pricing model, allowing for more precise estimation of the degree of diversion and changes in pavement deterioration rates. A number of important political considerations also surround the issue of roadway tolls, particularly on public highways, such as the issue of equity among travelers with different levels of income or wealth (Kockelman and Kalmanje, 2005; Duthie et al., 2007). Regulatory issues also play a major role—for instance, in the United States, Title 23 §301 of the United States Code generally prohibits highways constructed with federal money from tolls, with the exception of cases specified in §129(a), which would require federal negotiation.

Although these latter concerns are clearly important, most of the network pricing literature focuses primarily on the analytical and economic impacts of pricing. Likewise, the focus of this article is on the relationship between pavement deterioration, network assignment, and pricing, and equity and regulatory issues are beyond its scope. Rather, its contribution is determining tolls when the effects of diversion on both revenue and pavement deterioration are determined in an integrated manner, and on incorporating terms representing network infrastructure health into the objective function. Although the literature described in this section addresses one or two of these issues at a time (network diversion, revenue, and pavement deterioration), this article is the first to address all three simultaneously.

3 MODEL COMPONENTS

Virtually all pricing problems take the form of a Stackelberg game, with a “leader” and “followers” who each optimize distinct objective functions in sequence. For example, in a road pricing problem the leaders are the decision makers who set tolls, and the followers are the road users who aim to minimize their travel cost. The leader’s objective function, on the other hand, concerns broader social welfare, keeping in mind the followers’ reaction to any policy. A general formulation for this type of problem is

$$\max_{T, x} F(T, x)$$

subject to

$$T \in \Xi$$

$$x \in X$$

where $$x$$ represents the state of pavement infrastructure, specifically allowing toll revenue to fund additional
maintenance projects, in addition to user travel time-related costs. Users are divided into multiple classes based on value of time and weight. This classification is intended to represent heavy vehicles and passenger vehicles, but can represent a finer division as well. Further, the absence of congestion means that travel times can be well-approximated as constant. The following subsections give an overview of the model components, then describe each in turn: how drivers behave; how tolls are represented; how pavement M&R is modeled; and the criteria used by the agency in determining optimal tolls. Section 4 describes how these individual components are combined and lead to a solution algorithm.

### 3.1 Overview

The optimal tolling model consists of basic network assignment, revenue forecasting, pavement deterioration, and scheduling maintenance activities. This section provides a qualitative overview of each of these to clarify the overall structure of the model. The reader may find it helpful to refer to Table 1, which identifies all notation in the article classified as input parameters (that is, quantities that a practitioner must supply to run the model), outputs, and quantities calculated and used internally by the model.

From a high level, the model assumes three agents (the toll-setter, drivers, and the infrastructure itself), who make decisions in the following order. First, the toll-setter levies tolls on a subset of links. Second, drivers may change routes to minimize their generalized cost of travel. Third, as a result of this diversion, toll revenue and updated pavement deterioration rates can be calculated. Finally, an updated network maintenance policy can be adopted, considering the additional revenue from tolling and differences in pavement loading. The toll-setter aims to find an optimal set of tolls accounting for these factors, taking into account the “chain reaction” decisions of all the other agents just described. Specifically, the toll-setter’s objective consists of three components: benefits from improved maintenance due to toll revenue, changes in long-run maintenance costs due to diversion, and generalized cost of travel.

The network model assumes constant travel-time links, with discrete user classes who may have different values of time or vehicle axle loads. Tolls are assumed identical in both directions of a link. The pavement deterioration and maintenance model is disaggregate, using facility-level models for each link in the network, and solving a combinatorial knapsack problem to optimally allocate the budget across these links. Because the evaluation of each potential toll vector requires all of these models to be run in sequence, an efficient solution method is required to solve the model. The remainder of this section describes each model component in detail, whereas Section 4 describes a directed solution algorithm based on simulated annealing, and partitioning the feasible region based on the subset of links tolled.

#### 3.2 Driver behavior model

Consider a transportation network with sets of nodes \( N \), directed links \( A \), and zone centroids \( Z \) where trips can begin and end. In rural areas with no congestion, the travel time \( t_{ij} \) on each link \((i, j) \in A\) is a constant independent of flow. Assume that each link \((i, j)\) has a...
“mirror” link \((j, i)\) in the network. This assumption can be made without loss of generality: to simulate a one-way link, the travel time on the mirror can be set to a sufficiently large value.

There are \(K\) user classes in the model, each with a distinct value of time \(\beta^k\) and equivalent single-axle load \(L^k\}. These classes are specifically introduced to differentiate between heavy commercial vehicles and passenger vehicles. Although these are the two classes used in the demonstrations in Section 5, we present the model more generally (for instance, if the decision maker wishes to disaggregate heavy vehicles into multiple axle loads, or users into more value of time classes). Let \(d_{ij}^k\) denote the class \(k\) travel demand between centroids \(r\) and \(s\), and \(x_{ij}^k\) the total flow from class \(k\) on link \((i, j)\). If \(\Pi_{rs}\) is the set of all simple paths in \(G\) connecting \(r\) and \(s\), and \(h_{ij}^{k}\) is the flow on path \(\pi\) from class \(k\), then the set of feasible link flow vectors is

\[
X = \left\{ x_{ij}^k \in \mathbb{R}^{K |A|} : \begin{align*}
3h_{ij}^k &\geq 0 : \sum_{\pi \in \Pi_{rs}} h_{ij}^k = d_{ij}^k; \\
x_{ij}^k &= \sum_{(r, s) \in Z^2 \pi \in \Pi_{ri}(j) \in \pi} h_{ij}^k
\end{align*} \right\}
\]

(5)

The toll on each link \(T_{ij}^k\) is allowed to vary by user class, reflecting differential pricing. Each driver seeks to minimize their generalized cost of travel, defined by the total travel time on their chosen path multiplied by their class’s value of time, added to the total toll on that path. Because travel times are constant, this behavior is represented by setting the followers’ objective function (4) to

\[
f(T, x) = \sum_{(i, j) \in A} \sum_{k=1}^{K} (\beta^k x_{ij} + T_{ij}^k) x_{ij}^k
\]

(6)

This formulation assumes that travel costs and routing decisions are independent of pavement condition (see Steyn et al., 2011, 2012 for discussion of this issue). Although extending the model to include dependence on pavement condition is straightforward, doing so complicates the notation and is not undertaken here.

Notice that the demand values \(d_{rs}^k\) are not explicitly made elastic. In the demonstration that follows, diversion effects are captured by including a number of “external” nodes and links which represent trips diverting out of the network entirely. Incorporating a demand function in place of a constant trip table would not be difficult, but would complicate the model’s presentation and increase the computation time required. Further, with the long-distance nature of the travel involved in rural areas and high degree of commercial traffic (Schiffer, 2012), expensive tolls are more likely to lead to diversion rather than a decrease in total tripmaking. For these reasons, the model is based upon slightly enlarging the network to explicitly include potential diversion routes.

### 3.3 Tolling model

The vector \(T\) denotes the network tolls, with elements \(T_{ij}^k\) denoting the toll for vehicles of class \(k\) on link \((i, j)\). Any vector of nonnegative tolls is considered feasible if the toll on each link and its “mirror” is identical, that is, \(\Xi = \{\mathbb{R}^{K |A|} | T_{ij}^k = T_{ji}^k\} \in (2)\). This assumption is taken for two main reasons: first, from an implementation standpoint, it may be difficult to convince decision makers that charging different tolls in both directions is not arbitrary; and second, this assumption leads to improved solutions from the algorithm presented in Section 4. Although the model does not explicitly include political considerations, because including this particular factor improves the quality of the solutions found by the algorithm, this assumption is reasonable to take. Let \(\bar{R}\) be the gross revenue collected from all travelers, that is,

\[
\bar{R} = \sum_{k} \sum_{(i, j)} T_{ij}^k x_{ij}^k
\]

(7)

In practice, there are always some overhead costs associated with toll collection. Define the net revenue \(\Phi(\bar{R}, T, x)\), where \(\Phi\) reflects potential dependence of administrative costs on the gross revenue, toll vector, and flow vector. In general, all of these may be relevant: accounting and regulatory overhead can grow with the gross revenue \(\bar{R}\), the fixed cost associated with tolling a certain location (reflected by the positive elements of \(T\)), and the processing fees associated with each transaction (which depends on \(x\)). One possible function is

\[
\Phi(\bar{R}, T, x) = \bar{R} - \sum_{(i, j) \in A} \left( \xi + \sum_{k \in K} x_{ij}^k (\psi + \zeta T_{ij}^k) \right) \
\times I \left( \sum_{k \in K} T_{ij}^k > 0 \right)
\]

(8)

where \(\xi, \psi,\) and \(\zeta\) respectively denote the annualized cost for tolling a link, the per-vehicle overhead cost, and a coefficient representing income remaining after credit-card processing, and where \(I(\cdot)\) is an indicator function equal to one if its argument is true, and zero otherwise. This is the function used in the numerical analyses in Section 5, with parameters based on a Wyoming Department of Transportation toll feasibility study (2008a).
3.4 M&R model

The core contribution of this article is its integration of a network pricing and assignment model with a pavement deterioration, maintenance, and repair model. Although the overall model has network-level scope, to quantify the benefits of tolling more precisely, the model considers the state of individual facilities, which may exist in different condition and which will have different traffic loading and deterioration rates. Therefore, facility-specific models are introduced first, which will be incorporated into the network-level objective through the solution of a knapsack problem, discussed in Section 3.5.

To relate the traffic demand to the deterioration rate, let \( L_{ij} \) denote the total equivalent single-axle loading (ESAL) on link \((i, j)\), defined as the sum of loadings across all user classes:

\[
L_{ij} = \sum_k L_k^i x_{ij}^k
\]  \hspace{1cm} (9)

Further, an integral pavement condition rating \( PCR_{ij} \) is associated with each link, representing its quality (100 = like new, 0 = unusable). As described in Section 2, there are a great number of M&R models which have been proposed in the literature, but for reasons of computational efficiency an especially simple one is required—this model is a subroutine which must be run repeatedly. Thus, the starting point for the model is the relationship suggested by Mahoney and Jackson (1991), relating the pavement age \( \tau \) to its condition. This equation is modified below to account for the relationship between the actual loading \( L_{ij} \) and the design load \( L_{ij}^0 \).

There is a set of maintenance actions \( \Psi \) which can be performed in any given year. For instance, these may include routine maintenance or complete replacement. Each action \( a \in \Psi \) increases a link’s \( PCR \) by a fixed improvement value \( E_a \), with a cost of \( C_a \). A always includes the “do nothing” action with no cost or effect.

The transition function \( \phi(PCR, L, a) \) maps a link’s current \( PCR \) value to the next year’s \( PCR \) value, given the ESAL loading \( L \) and the maintenance action \( a \) applied. Equation (10) is transformed by scaling the pavement age \( \tau \) by \( L_{ij}/L_{ij}^0 \), the ratio between actual and design loadings. Specifically, given the \( PCR \) value, the pavement age can be obtained by inverting (10), giving

\[
\tau = \left( \frac{100 - PCR}{0.76} \right)^{1/1.75}
\]  \hspace{1cm} (10)

while the effective age in the next year is given by \( \tau + \frac{L_{ij}}{L_{ij}^0} \). Substituting this into (10), adding the effect of any maintenance action, and rounding down to integer values yields

\[
\phi(PCR, L, a) = \min \{100, \left[ 100 - 0.76 \left( \frac{100 - PCR}{0.76} \right)^{1/1.75} + \frac{L_{ij}}{L_{ij}^0} \right] + E_a \} \hspace{1cm} (12)
\]

Any action \( a \) for which \( \phi(PCR, L, a) < 0 \) is deemed infeasible.

With this transition function, the optimal maintenance actions to perform in any year can be found through dynamic programming. Letting \( Y \) represent the time horizon, and \( y \) any year between the present and \( Y \), one can identify the maintenance actions and, more importantly for this article, a value function \( V(PCR, L, 0) \) indicating the economic value of a facility with condition \( PCR \), including expected future maintenance outlays and a terminal “salvage value” \( V_0(PCR) \) for each link based on its \( PCR \) at year \( Y \). Introducing a discount factor \( \alpha \) ensures proper accounting of future costs. The value function can be calculated by solving the backward recursion

\[
V(PCR, L, Y) = V_0(PCR)
\]  \hspace{1cm} (13)

\[
\arg \max_a \{(1 - \alpha)V(\phi(PCR, L, a), L, y + 1) - C_a\}
\]  \hspace{1cm} (14)

\[
V(PCR, L, y) = (1 - \alpha)V(\phi(PCR, L, a^*(PCR, L, y)), L, y + 1) - C_a
\]  \hspace{1cm} (15)

solving (14) and (15) for \( y \in [Y - 1, Y - 2, \ldots, 0] \). These value functions are used to quantify the benefit of maintenance actions and calculate the present value of forecasted long-term expenditures.

3.5 Leader’s objective function

This section describes how the leader’s objective function \( F(T, x) \) in (1) is formulated. In the model, the toll-setting agency has three considerations in choosing tolls: (1) the benefits that toll revenue provides in funding additional maintenance projects; (2) the impact on long-term maintenance costs and economic pavement value caused by traffic diversion; and (3) the impact of tolls on the generalized cost of travel (for both local and pass-through traffic). By converting all of these to cost units,
these three considerations become commensurate, and the leader chooses tolls to optimize a linear combination of these quantities. Each of these is described in turn.

The value of the benefits from additional revenue is denoted $B(T, x)$, expressing the dependence of the revenue on both the toll vector $T$ and flow vector $x$. These benefits are calculated by enumerating all possible maintenance projects, their costs, and their benefits, and then solving a knapsack problem to maximize benefits given the amount of toll revenue available.

The number of potential projects is the product of the number of links and the size of $\Psi$ (omitting the “do-nothing” action). The benefit of performing action $a$ on link $(i, j)$ is the difference between the value function after performing action, and the current value function:

$$B_{ij}^a = V(\phi(PCR_{ij}, L_{ij}(x), a), L_{ij}(x), 0) - V(PCR_{ij}, L_{ij}(x), 0)$$

whereas the cost of this project is $C_{ij}^a(x) = C_a$.

The problem of finding the most beneficial set of projects allowed by the budget is an instance of the knapsack problem, a standard combinatorial optimization problem (see, for instance, Silvano and Toth, 1990). This knapsack problem uses the benefits $B_{ij}^a$, costs $C_{ij}^a$, and a budget of $R$, the net revenues. The total benefits $B(T, x)$ are the value of the optimal solution to this knapsack problem. (Solution methods are discussed in Section 4.)

The change in link flows from diversion also influences the long-range infrastructure maintenance costs and economic value. In particular, one expects maintenance costs to decrease on tolled links, and increase on links used as alternate routes, based on shifting traffic patterns. These impacts are quantified in the infrastructure present value $M(x)$, defined as

$$M(x) = \sum_{(i, j) \in A} V(PCR_{ij}, L_{ij}(x), 0)$$

Finally, the total generalized cost of travel is simply $f(T, x)$ from (6). Using the quantities defined above, the leader’s objective function (1) is to maximize

$$F(T, x) = B(T, x) + M(x) - f(T, x)$$

recognizing that $B$ and $M$ denote benefits, whereas $f$ denotes a cost.

4 SOLUTION ALGORITHM

As described in the previous section, the rural pricing model contains several interconnected submodels representing pavement deterioration and maintenance, and multiclass driver route choice. This section describes how to choose tolls to optimize the objective function $F$ specified in (18). Section 4.1 first describes how $F$ is calculated for a given toll vector $T$, allowing $F$ to be treated as a “black box” in the simulated annealing heuristic described in Section 4.2. Section 4.3 describes the complete solution algorithm which involves repeated application of simulated annealing on a partitioning of the feasible region.

4.1 Calculating $F$

Given a particular toll vector $T$, the followers’ objective function (6) completely determines the values of $x$; therefore, the leader’s objective (18) can be parameterized in $T$ alone, that is, as $F(T) = F(T, x(T))$. The value of this function can then be calculated by performing the following steps in order:

1. For each vehicle class $k$, calculate shortest paths with respect to the generalized costs $b^k_i + T^k$ using a standard algorithm (see, e.g., Ahuja et al., 1993). Then, for each OD pair $(r, s)$ and class $k$, load $d_{rs}^k$ vehicles onto the relevant shortest path to obtain the class-specific link flows $x^k$. This gives the total generalized cost of travel $f(T, x)$ via (6).
2. Determine the ESAL loading $L_{ij}$ on each link $(i, j)$ by applying (8), based on the calculated link flows.
3. Calculate gross revenue $R$ using (7).
4. Calculate net revenue $R = N(R, T, x)$, accounting for administrative costs.
5. Recalculate the infrastructure value $M(x)$ using the updated ESAL loadings, thereby accounting for changes in deterioration rates due to diversion and the associated long-term maintenance costs. This involves solution of the backward recursion (13), (14), and (15) as a standard dynamic programming problem (see, for instance, Bertsekas, 2005, 2007).
6. Identify all candidate maintenance projects and their associated benefits, applying (16).
7. Identify the best set of candidate projects, maximizing the benefits calculated in the previous step while respecting a budget based on the net revenue $R$. This involves solution of a knapsack problem, and the value of the optimal knapsack gives the value of $B(T, x)$.
8. Finally, calculate $F(T, x) = B(T, x) + M(x) - f(T, x)$ with all three values on the right-hand side determined in the steps above.

Notice that all of these steps must be performed whenever the solution method evaluates a potential toll vector $T$, which may occur thousands of times or more.
Therefore, the algorithms and implementations used in these steps should be as efficient as possible. In the demonstrations which follow, the shortest path algorithm in Step 1 was solved using the Bellman-Ford algorithm with a first-in-first-out scan list (Bellman, 1958). The dynamic programming problem described in Step 5 can have a rather large state space, depending on the discretization used. To reduce the computational burden, the value functions can be calculated once as a preprocessing step, with all possible values stored in a lookup table. Although this method increases the memory requirements, the computation time can be substantially decreased. The knapsack problem in Step 7 was solved by applying a cost discretization, followed by a standard dynamic programming algorithm described in Silvano and Toth (1990).

4.2 Simulated annealing

The objective function $F$ is nonconvex, nondifferentiable, and discontinuous, and therefore lacks the regularity properties required by exact solution methods. Therefore, heuristic methods must be applied. As reported in Section 5, a simulated annealing-based algorithm provides acceptable results in a reasonable amount of time. Simulated annealing is a general-purpose metaheuristic, first described in Kirkpatrick et al. (1983), which requires no regularity conditions on the objective function. Instead, simulated annealing only requires the ability to evaluate the objective function for a feasible set of decision variables, the description of a “neighborhood,” and a “cooling schedule” of algorithm parameters. The former was discussed in the previous subsection; the latter two components are described next.

Starting from an initial feasible solution, simulated annealing attempts to find an optimal solution by iteratively moving from one feasible solution to another. At any stage, the set of potential candidate moves defines the neighborhood. After experimenting with several neighborhood choices, the following definition was found to provide the best performance in the overall algorithm described in Section 4.3: given a current toll vector $T$, the neighborhood $\Omega(T)$ consists of all toll vectors $\hat{T}$ which (1) are themselves feasible; (2) do not deviate more than a pre-specified amount $\varepsilon$ in the toll value for any user class on any link; and (3) require that the links tolled in $T$ are exactly those tolled in $\hat{T}$. That is,

$$\Omega(T) = \times_{k=1}^{K} \left\{ \hat{T} \in \mathbb{N}^{[A]} \mid \begin{array}{l} [\hat{T}^{k}]_{ij} = [\hat{T}^{k}]_{ij} \\ |T^{k}_{ij} - \hat{T}^{k}_{ij}| \leq \varepsilon \quad \forall (i, j) \in A \\ T^{k}_{ij} = 0 \Rightarrow \hat{T}^{k}_{ij} = 0 \end{array} \right\}$$

(19)

This third requirement may seem unusual, but it plays a key role in the larger solution algorithm discussed in the following subsection, and an explanation is provided there.

Given an incumbent solution $T$, simulated annealing proceeds by randomly generating a neighboring solution $\hat{T} \in \Omega(T)$ and comparing the objective function values $F(T)$ and $F(\hat{T})$. If $F(\hat{T}) \geq F(T)$, the new tolls are superior to the incumbent tolls, $\hat{T}$ is adopted as the new incumbent solution, and the process repeats. On the other hand, if $F(\hat{T}) < F(T)$, $\hat{T}$ is only adopted as the incumbent solution with some probability $p$, with no change in the incumbent solution otherwise. By allowing the possibility of a nonimproving move, simulated annealing is thus capable of escaping from local optima that a purely local search can become trapped in. The “cooling schedule” describes how the threshold probabilities $p$ change during the solution process. The typical rule $p = \exp((F(\hat{T}) - F(T))/\theta)$ is adopted, where $\theta$ is the current “temperature” dictated by the cooling schedule—thus, the likelihood of moving to a disimproving solution depends both on how much worse the new solution is, and on the current temperature. The initial temperature $\theta_{0}$ is high, suggesting that most moves will be accepted. As the algorithm progresses, the temperature is periodically decreased until termination.

Specifically, the simulated annealing algorithm works as follows:

1. Choose an initial feasible toll configuration $T$, and calculate its objective value $F(T)$ as described in Section 4.1.
2. Initialize the best tolls found so far $T^{*} \leftarrow T$ and best objective function value $F^{*} \leftarrow F(T)$.
3. Randomly generate a candidate solution $\hat{T} \in \Omega(T)$ and calculate $F(\hat{T})$.
4. If $F(\hat{T}) > F(T)$, update $T^{*} \leftarrow \hat{T}$ and $F^{*} \leftarrow F(\hat{T})$.
5. If $F(\hat{T}) \geq F(T)$, move to the candidate solution: $T \leftarrow \hat{T}$.
6. If $F(\hat{T}) < F(T)$, move to the candidate solution with probability $\exp((F(\hat{T}) - F(T))/\theta)$.
7. Update $\theta$ according to the cooling schedule, and return to Step 3 unless termination criteria are met.
8. Return $T^{*}$ as the optimal tolls.

The cooling schedule is chosen in a problem-specific manner, using a procedure adapted from Chiang and Russell (1996). The initial temperature $\theta_{0}$ is chosen based on the initial feasible toll configuration, so that the probability of accepting a disimproving move is 0.05 (based on 4,000 random samples of the initial neighborhood). This temperature is maintained constant for a number of iterations (20 times the number of links tolled), before being decreased by 5%. These specific values
were found to give good results with this model, based on preliminary tests measuring the objective functions attained over the 1-hour computational budgets. Simulated annealing terminates when the proportion of “accepted” candidate solutions drops below 1%.

4.3 Finding optimal tolls

It is possible to apply simulated annealing to the entire feasible region \( \mathcal{Z} \) at once, omitting the third requirement in the neighborhood definition \( (T_i^j = 0 \Rightarrow \hat{T}_i^j = 0) \) for all \((i, j) \in A\). In this way, the algorithm is free to adjust tolls on all links in any move. However, experiments showed that a more directed version of the algorithm provides superior solutions in less time. The presence of administrative costs suggests that optimal solutions will toll relatively few links. Furthermore, in practice, engineering judgment may suggest that certain links are more likely candidates for tolling. The directed solution method exploits these features of the rural pricing problem by partitioning the feasible region in a way that allows prioritization of certain tolled link configurations.

Consider some subset of links \( A \subseteq A \), and let \( \mathcal{Z}(\hat{A}) \) be the set of feasible tolls in which only links in \( \hat{A} \) are tolled, that is,

\[
\mathcal{Z}(\hat{A}) = \left\{ T \in \mathcal{Z}^{k \mid A} \mid T \in \mathcal{Z}; \quad T_{ij}^k = 0 \forall (i, j) \notin \hat{A} \right\}
\]  

Note that each of these sets is closed under the neighborhood definition (19) in the sense that if \( T \in \mathcal{Z}(\hat{A}) \), then \( \Omega(T) \subseteq \mathcal{Z}(\hat{A}) \), and that collectively the sets \( \mathcal{Z}(\hat{A}) \) partition \( \mathcal{Z} \). Thus, if simulated annealing is initiated with an element of \( \mathcal{Z}(\hat{A}) \), all generated solutions will remain in the same set. This suggests a “partitioning” algorithm, where simulated annealing is run sequentially, starting with a feasible solution from each of the sets \( \mathcal{Z}(\hat{A}) \). In this way, the best solutions from each subset are identified, with each search requiring much less time because it is focused on a smaller, more structured subset of the feasible space.

Because the computational budget likely prohibits complete enumeration of \( \hat{A} \), the subsets should be prioritized in some fashion. As solutions tolling few links are likely to be preferable (both due to administrative costs and due to political feasibility), the implementation first iterates over all subsets of size one, then all subsets of size two, and so forth. If there are a priori reasons to favor including or excluding any links from consideration from tolling, these can easily be accommodated in the construction of these subsets.

In general, the algorithm consists of the following steps:

1. Initialize the best tolls found so far with the “do-nothing” option: \( T^{**} = 0 \), \( F^{**} = F(0) \).
2. Choose an initial subset of tolled links \( \hat{A} \).
3. Generate a feasible solution \( T^0 \in \mathcal{Z}(\hat{A}) \).
4. Perform simulated annealing with \( T^0 \) as the feasible solution, obtaining \( T^* \) and \( F^* \) as the best tolls and objective function value for the subset \( \hat{A} \).
5. If \( F^* > F^{**} \), update the best solution found so far: \( T^{**} \leftarrow T^* \), \( F^{**} \leftarrow F^* \).
6. Generate the next subset of tolled links \( \hat{A} \).
7. Return to Step 2 unless termination criteria are met.

In the implementation used to generate the results in the following section, the sole termination criterion was a limit of 1 hour of computation time, and subsets are generated in order of cardinality as suggested in the previous paragraph.

5 NUMERICAL ANALYSIS

This section demonstrates this pricing model on a network representing the state of Wyoming in the United States. The population of Wyoming is the least of any state in the country (563,626 as of the 2010 census), making the vast majority of the state rural. The main east–west freeway is I-80, a major freight corridor linking San Francisco to New York City, for which reasonable alternative routes are lacking. The Wyoming Department of Transportation also recently completed a feasibility study on tolling I-80 (Wyoming Department of Transportation, 2008b). For these reasons, Wyoming is an ideal testbed for the model described here.

Figure 1 shows a diagram of the Wyoming network, which contains 28 zones, 60 nodes, and 188 links. In this
Two user classes were considered, passenger cars and semitrucks. The ESAL equivalencies for these two user classes are taken as 0.0007 and 0.39, using typical values (American Association of State Highway and Transportation Officials, 1993). Values of travel time for these two user classes are taken to be $10/hour and $60/hour, respectively. Trip tables for each user class were estimated using a least-squares method based on volume counts published by the Wyoming Department of Transportation (Wyoming Department of Transportation, 2008a) and a gravity trip distribution model, a method described in Bell and Iida (1997). Link travel times are estimated as the quotient of link length and the speed limit. Two maintenance actions are considered, routine maintenance which increases PCR by 10 at a cost of $1,000,000 per mile, and replacement which restores PCR to 100 at a cost of $5,000,000 per mile. To represent administrative overhead, the parameters $\xi$, $\psi$, and $\zeta$ in Equation (8) respectively take the values of $860,000$ (the annualized cost for tolling a link, based on amortizing a $12 million total estimated expenditure over a 30-year life span at a 6% discount rate), $0.27$ (the per-vehicle overhead cost), and 0.97 (reflecting a 3% loss in credit-card processing fees). All of these values were obtained from a technical report on toll feasibility in Wyoming (Wyoming Department of Transportation, 2008b). A 60-year time horizon was used for calculating long-range maintenance costs.

Figure 2 shows the optimal solution, with a recommended toll of $23.20 for passenger vehicles and $71.81 for heavy vehicles on I-80 between Rock Springs, WY and Rawlins, WY. These tolls are applied at the same location as in the profit-maximizing solution recommended in the initial feasibility study, but take different values. The model suggests a considerably lower toll for heavy vehicles ($71.81 as opposed to $116), but somewhat higher for passenger cars ($23.20 compared to $9.46). This solution generated an estimated net annual revenue of $248 million, which was spent on roadway maintenance projects with a total benefit-cost ratio of 4.89. Annual average daily traffic (AADT) at the toll location decreased from 8,859 passenger vehicles and 3,628 heavy vehicles with no tolls, to 7,087 passenger vehicles and 3,265 heavy vehicles with these tolls.

The remaining experiments include sensitivity analysis on three model parameters: the trip table, the initial state of the roadway pavement, and the ESAL factor for heavy vehicles. To change the trip table, all values were adjusted by a demand multiplier ranging from 0.5 to 2.0, capturing sensitivity of the model to the accuracy of these parameters, and providing guidance for future years when demand is likely higher. (See Duthie et al., 2011, for more on the importance of multiple demand scenarios in transportation planning.) Figure 3 compares the toll and no-toll values for the objective...
function as the demand multiplier varies. Two observations are worth noting: first, the objective function decreases with demand even in the presence of tolls, indicating that the impact of higher roadway volume on pavement state cannot be completely compensated for by tolling. Second, the benefits of tolling increase with demand. Both of these results are intuitive.

Figure 4 plots the annual revenue according to the demand multiplier; this relation is increasing and approximately concave. Figure 5 shows the benefit-cost ratio of the projects funded by toll revenues; a roughly increasing trend is seen here with respect to demand, even as Figure 3 showed a decreasing objective function with demand. This may be related to two competing effects: as demand increases, revenue increases and more expensive (yet highly beneficial) projects can be funded; yet at the same time, the higher deterioration rate associated with higher demand increases long-term maintenance costs, which are penalized in a different term in the objective function. Figure 6 shows the locations selected for tolling, along with the number of demand scenarios in which that location was chosen. For every scenario except the lowest demand, I-80 was tolled between Rock Springs and Rawlins, likely due to a combination of high volume and limited diversion opportunities—in fact, for every demand scenario at least 20% above the current trip table, this was the only location chosen for tolling. The stability of the optimal tolling location with respect to demand suggests that tolling at this location is likely to remain optimal even into the future as demand increases.

Figure 7 shows the change in passenger car and heavy vehicle AADT on I-80 between Rock Springs and Rawlins by demand multiplier. In most cases, the toll is high enough to deter 9% of traffic. Due to the
network topology, all diverting vehicles at this location do not travel through Wyoming at all, but on I-70 through Colorado (the bottommost external link in Figure 1). At this point, the benefits from revenue collected, the change in pavement deterioration, and the additional cost to drivers are approximately in equilibrium. The “spikes” in AADT change for particular demand multipliers suggest that multiple optimal solutions may exist, as is common when objective functions lack regularity properties such as convexity—there may be two optimal (or nearly optimal) solutions with roughly similar objective function values, but with substantially different toll amounts. For demand multipliers of 0.8 and 1.2, the solution involved a higher toll (and thus greater diversion), which is somewhat evident in Figure 4, where the dip at a demand multiplier of 1.2 is consistent with the large diversion seen in Figure 7. However, the smooth trend in the leader’s objective $F$ (Figure 3) suggests that the overall trade-offs between deterioration rates, revenue, and user cost remain the same across demand levels, despite fluctuations in the components of the objective returned by the heuristic.

A second sensitivity analysis studies how the solution varies according to the initial state of the pavement infrastructure, and is plotted in Figure 8. When the initial pavement state is worse, the revenue collected in the optimal solution is much higher, indicating that substantially higher tolls are needed to offset pavement deterioration. This is intuitive, given the concave deterioration relationship used in this research—pavement in worse condition deteriorates at a faster rate. Contrariwise, if the pavement infrastructure is initially in healthy condition, much lower tolls are needed.

As a final sensitivity analysis, Figure 9 shows how maintenance expenses at the optimal solution vary based on the ESAL value of the heavy vehicles. This particular sensitivity analysis is important because the rapid growth in hydraulic fracturing in rural areas has led to a very different load profile than was assumed during the design process. As shown in this figure, there is a nearly linear relationship between the logarithm of the ESAL value for heavy vehicles and the logarithm of annualized maintenance expenditures.

The convergence profile of the “base case” is shown in Figure 10, showing the objective function values corresponding to the best solution found over the hour of computation time allotted. There is no improvement in the best solution found after approximately 10 minutes, suggesting that the solution method has likely found the best solution it is able to find, and that the hour of computational time is a conservative choice.
The purpose of these experiments is not to draw definitive conclusions about recommended toll values and revenue forecasts for Wyoming, but to illustrate the operation of the model developed here, and its utility in developing pricing policies in rural areas. The recommended values for heavy vehicles are significantly lower than those found in Wyoming Department of Transportation’s feasibility study; this is logical, as the goal of the model in this article is maximizing total benefits to society, rather than revenue maximization.

6 CONCLUSIONS

This article describes a bilevel programming model to solve the rural toll pricing problem. The leader wants to find tolls maximizing an objective function containing terms representing infrastructure quality and expense to travelers, whereas the followers choose routes to minimize travel cost. The central contribution of the model is the integration of a pavement deterioration and maintenance model into toll selection. In particular, the relationship in Equation (12) was derived to provide a reasonable, yet highly efficient, deterioration model. Simulated annealing was applied to find tolls for the Wyoming network as a case study. Further, a new solution method has been devised, based on partitioning the feasible region and applying simulated annealing. From the standpoint of practice, this model includes a broad set of factors (revenue, diversion, and pavement deterioration) within a single, unified modeling and optimization framework. To the authors’ knowledge, such a framework has not been available for rural roadway tolling. This model has been designed with simplicity in mind, and required data (Table 1) are all relatively easy to obtain.

There are several important ways to improve the scope of this model. Especially in rural regions, freeway traffic provides a major source of economic activity in small towns. Accounting for the impact of diversion on economic activity in these towns would improve the recommendations of the model. Applying the extensions of the basic congestion pricing problem discussed in Section 2, such as a continuous distribution of user classes, elastic demand, or stochasticity in demand or roadway conditions, would be helpful as well. Replacing the simplistic pavement deterioration model by a more sophisticated one can enhance the accuracy of the model’s predictions, but the impact on computation time would need to be carefully studied. The model and solution algorithm are designed with modularity in mind: for instance, to use a more sophisticated pavement model, the only major changes would be to Steps 5–7 in the algorithm in Section 4.1. To use a more sophisticated network assignment model, such as one allowing for congestion in route choice, the only change would be to Step 1 (in this case, replacing the lower-level objective (6) with the Beckmann function and solving for user equilibrium, cf. Sheffi, 1985). The solution method and remaining components could still be used as is.

From the standpoint of practical implementation, it would be interesting to compare the findings from the Wyoming network to those in other rural networks, with different demand patterns—for instance, some rural areas are currently experiencing population declines, which will affect future benefits and costs. Finally, incorporating equity and other political factors, such as those described in Section 2, is also of great practical importance.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Mountain-Plains Consortium in sponsoring this research (Project #320). Rebecca Franke provided valuable assistance with the Wyoming data and model runs. Nine referees and the editor also provided helpful suggestions to improve the article. The conclusions are the authors’ alone, and do not reflect positions of the Wyoming Department of Transportation, the Mountain-Plains Consortium, or any other organization.

REFERENCES

Structures, American Association of State Highway and Transportation Officials, Washington DC.


Schiffer, R. G. (2012), Long-Distance and Rural Travel Transferable Parameters for Statewide Travel Forecasting Models. NCHRP Report 735, Transportation Research Board, Washington DC.


Wyoming Department of Transportation (2008b), Interstate 80 Tolling Feasibility Study, prepared by Parsons Brinckerhoff in association with PB Consult, New York, NY.

