

Optimal Locations of Railroad Wayside Defect Detection Installations

Yanfeng Ouyang,* Xiaopeng Li, Christopher P. L. Barkan, Athaphon Kawprasert

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA

&

Yung-Cheng Lai

Department of Civil Engineering, National Taiwan University, Taipei, Taiwan

Abstract: Railroads have been using wayside inspection technologies for many years to improve the safety and efficiency of operations. There has been a recent proliferation of new, more sophisticated, but also more costly inspection systems that are capable of detecting a wide range of subtle defects before a problem has actually occurred. This new class of “predictive” wayside detection systems, as compared with the older “reactive” technologies, requires a different deployment strategy. The higher cost of these new technologies means that it is particularly important for railroads to maximize the benefit they derive from the investment. This article presents a network optimization model that selects cost-effective installation sites for wayside defect detection systems over a railroad network. The objective is to maximize the total inspection benefits possible under any given investment budget. We develop solution techniques based on Lagrangian relaxation to effectively solve the problem. The article also presents case studies with empirical data to illustrate the technique. The computational results show that the problem can be solved efficiently, and that the model has the capability of being applied to full-scale railroad networks at regional or national levels. There are a variety of ways that railroads can use this model to help them more efficiently invest in wayside inspection

technology so as to maximize the safety and economic benefits of these technologies.

1 INTRODUCTION

Railroads use various different types of technologies installed adjacent to railroad tracks to monitor the health and performance of passing rolling stock. The design and technology used in these “wayside detectors” vary widely depending on what part or characteristic of the railcar’s performance or condition they inspect. The technological maturity of these systems also varies. For example, dragging equipment detectors were first developed in the 1930s (Post, 1936, 1937; Burpee, 1945), and wayside hot bearing detectors were developed in the 1940s and 1950s (Austin, 1949; Gallagher and Pelino, 1959), whereas others are still in the testing or early deployment stage (Blevins et al., 2003; Resor and Zarembski, 2004; Barke and Chiu, 2005; Lundgren and Killian, 2005; Lagneback, 2007).

Several factors affect the strategy regarding the number and location of various system installations. These include the problem or aspect of the car’s condition the particular technology is designed for, the consequences if the problem is not detected, particularly the time between when a problem becomes detectable and when

*To whom correspondence should be addressed. E-mail: yfouyang@uiuc.edu.

an action needs to be taken in response. In this context, Lagnebäck (2007) distinguishes between “reactive” and “predictive” detectors. Reactive systems detect faults on vehicles that may have short latency between fault and catastrophic failure (e.g., derailment). Dragging equipment and hot bearing detectors are the widely used examples of reactive technologies. The strategy with regard to deployment of such systems is to place them along tracks where the traffic density and operational practices justify them. The detectors are typically spaced closely enough (e.g., every 20 miles) so that a failing bearing can be detected in time to take remedial actions (English, 1996; CN, 2008). There is no advance warning until fairly severe remedial action is required right away.

Predictive systems on the other hand are capable of measuring, recording, and trending the performance of the vehicles and specific components. The information collected can be used to analyze the condition of equipment to predict possible failures and faults that may occur some time in the future, thereby making it easier to plan maintenance activities and use and repair equipment more efficiently (Blevins et al., 2003; Tournay and Cummings, 2005; Tournay et al., 2006; Lagnebäck 2007). Such systems have become increasingly feasible and cost-effective over the past two decades with the rapid advancement in sensor, information-processing, and communications technology (Steets and Tse, 1998; Lundgren and Killian, 2005). An example is the expanding installation of truck performance detectors (TPD) that provide various diagnostic information on the health of the truck and related components (Wolf and Peterson, 1998; Tournay et al., 2006). TPDs use an array of strain gauges mounted on the rails to measure vertical and lateral loads on the track structure as railcars pass through a particular type of track configuration. They integrate the temporal and spatial patterns of these loads so that certain patterns indicative of the condition of various elements of the wheels and truck (bogie) assembly can be assessed. Another example is the Association of American Railroads’ (AAR) Fully Automated Car Train Inspection System (FACTIS), which is currently undergoing field testing by the AAR and railroads (Lundgren, 2007). Other advanced inspection technologies are also in the R&D stage, such as machine vision systems to detect the loading efficiency of intermodal trains (Lai et al., 2007) or to inspect safety appliances on railcars (Edwards et al., 2006).

The greater sophistication and capability of these systems come at a considerably higher cost. The cost to install and maintain the technology and the benefits it provides are important factors affecting deployment (Resor and Zarembski, 2004; Barke and Chiu, 2005).

Each installation may be on the order of three quarters of a million dollars, compared with \$100,000 (or less) for more conventional systems such as hot bearing or dragging equipment detectors. Furthermore, the underlying philosophy of these advanced technologies is fundamentally different (Lagnebäck, 2007). As they are not detecting imminent failure, deploying large numbers of them throughout the network is not as important as is maximizing the likelihood that the largest possible number of distinct cars will be “seen” on a regular, but not necessarily frequent, basis. Consequently, fewer of these detectors are needed compared with conventional detector systems, and the characteristics of the deployment locations also may differ. Railroads are among the most capital intensive of all industries, and thus they are acutely sensitive to the cost of equipment such as this. As the demand for deployment of these technologies increases, the importance of knowing how to locate them to maximize the safety benefits they provide becomes ever more important. There needs to be more consideration of where they should be placed to maximize the utility they provide to railroads in the most economically efficient manner possible.

North American railroads are made up of a complex system of intersecting lines and routes; approximately 1.3 million freight cars in the North American fleet make over 31 million trips over portions of this network each year (AAR, 2006). Understanding how to best locate wayside inspection systems is fundamentally a network-level problem. Although certain cars such as those in unit or intermodal train service tend to travel together over fairly confined routes, a substantial number of cars involved in carload freight service travel much more widely across the network. Ensuring that all of these cars receive inspections on a regular basis is challenging.

There are a variety of criteria railroads can and do use to select locations to install detectors, but we are unaware of any comprehensive, formal quantitative framework that allows them to determine the optimal placement of predictive wayside inspection facilities. For many of the new predictive technologies, it may be adequate to inspect a particular car only infrequently, for example, every 6 months or so, with the objective to maximize the number of distinct cars inspected. If on a given route, most of the traffic traverses its entire length, then multiple installations along that route provide almost no additional value. Instead, inspection facilities should be placed at other locations along other routes with different traffic and routing patterns.

The development of an optimization approach for the placement of wayside inspection facilities that takes all

cars and routings into account is nontrivial and computationally intensive. The field of optimization modeling and operations research offers the tools to solve this problem. Furthermore, the increasing computational power now widely available means that complex algorithms using very large amounts of data can be executed in a feasible time frame, thereby providing railroads with better information on how to optimize their installation strategies for new inspection technologies. In this article, we develop a quantitative framework to address the problem of optimal placement of wayside rolling stock inspection facilities. The methodology is adapted from several well-known network design and optimization techniques that have proven useful for solving similar problems. The model is designed to be flexible regarding the types of questions it can address, the weight assigned to different criteria that affect the outcome, and the type of data available to use the model.

The exposition of the article is as follows. Section 2 introduces the mathematical formulation and a solution algorithm based on Lagrangian relaxation (LR). Section 3 presents numerical studies with hypothetical and empirical data. Section 4 provides conclusions and future research.

2 METHODOLOGY

2.1 Formulation

To systematically select optimal locations, we formulate the problem as follows. Let I be the set of all candidate installation locations. Each candidate location is indicated by $i \in I$, where we can potentially install one and only one of M facility types¹. Installing facility type $m = 1, 2, \dots, M$ at candidate location i would incur a characteristic cost (e.g., equipment acquisition and supply and maintenance) of c_{im} ; if facility type m cannot be installed at location i , for whatever reason, the installation cost can be set to $c_{im} = \infty$, so that the model will reject such an option. Overall, the installation costs for all facilities at all locations cannot exceed a total budget of B .

Railcars having the same origin-destination pairs and travel paths are considered to be part of the same car flow. Let J be the set of railcar flows. For any flow $j \in J$, suppose the volume (i.e., the number of cars) is h_j . We let $a_{ij} = 1$, if flow $j \in J$ passes candidate location $i \in I$, or $a_{ij} = 0$ otherwise. Obviously, railcar flow j will be inspected for potential defects at location i if $a_{ij} = 1$ and a facility exists at location i . Suppose too that each railcar in flow j will receive an economical benefit of f_j if inspected correctly.

Facility types differ by their technology specifications and capability of detecting railcar defects. Assume that a type- m facility is able to correctly inspect a passing railcar (i.e., provide accurate inspection results) with probability $(1 - q_m)$, no matter how many times that railcar is inspected by that type of facility. Then, we let $k_m = 1, \forall m = 1, 2, \dots, M$, if a railcar flow passes a type- m facility at least once, or $k_m = 0$ otherwise. Then vector $\mathbf{k} = [k_1, k_2, \dots, k_M] \in \{0, 1\}^M$ fully specifies how a railcar flow is inspected by the M types of facilities. For example, a trivial scenario is $\mathbf{k} = \mathbf{0}$, indicating that a flow is not inspected at all. The total number of such inspection scenarios is $P = 2^M$. It is also reasonable to assume that the probability for one facility type to inspect one car correctly is independent of the probability for another facility type or another car. Then, when a railcar flow is inspected by a combination of at least one facility (i.e., $\mathbf{k} \neq \mathbf{0}$), the probability that any car inside this flow will be correctly inspected is $1 - \prod_m q_m^{k_m}$, and the expected benefit for inspecting flow j in scenario \mathbf{k} is estimated as

$$b_{j\mathbf{k}} = h_j f_j \left(1 - \prod_m q_m^{k_m} \right), \quad \forall j, \mathbf{k}. \quad (1)$$

The inspection benefits are input data to the optimization model that we will discuss next. Computation of the benefits, that is, the structure of Equation (1), can be very flexible. Under different real-world situations, different formulas may be applied without affecting the optimization model and solution procedure. For example, Equation (1) only considers detection benefits from correctly identifying defects; however, other factors (e.g., false alarm rates) can also be incorporated. The benefit may further include diagnosis of different types of problems to different extents. In some other cases, we may allow k_m to be an integer that exactly equals the number of times a car is inspected by a type- m facility.

We define two sets of variables, $\mathbf{x} = \{x_{im}, \forall i, m\}$ and $\mathbf{y} = \{y_{j\mathbf{k}}, \forall j, \mathbf{k}\}$, where

$$x_{im} = \begin{cases} 1, & \text{if facility type } m \text{ is installed at location } i \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, m$$

and

$$y_{j\mathbf{k}} = \begin{cases} 1, & \text{if flow } j \text{ is inspected in scenario } \mathbf{k} \\ 0, & \text{otherwise} \end{cases}, \quad \forall j, \mathbf{k}$$

We can now formulate a mathematical program that maximizes the total expected inspection benefits subject

to the budget constraint, that is

$$\text{Max}_{\mathbf{x}, \mathbf{y}} \sum_j \sum_{\mathbf{k}} b_{j\mathbf{k}} y_{j\mathbf{k}} \quad (2a)$$

$$\text{s.t.} \sum_i \sum_m c_{im} x_{im} \leq B \quad (2b)$$

$$\sum_m x_{im} \leq 1, \quad \forall i \quad (2c)$$

$$\sum_{\forall \mathbf{k} \in \{0,1\}^M} k_m y_{j\mathbf{k}} \leq \sum_i a_{ij} x_{im}, \quad \forall j, m \quad (2d)$$

$$\sum_{\forall \mathbf{k} \in \{0,1\}^M} y_{j\mathbf{k}} = 1, \quad \forall j \quad (2e)$$

$$x_{im}, y_{j\mathbf{k}} \in \{0, 1\}, \quad \forall i, m, j, \mathbf{k} \quad (2f)$$

The objective function in Equation (2a) maximizes the expected inspection benefits; the benefit coefficient is given by Equation (1). Constraint in Equation (2b) imposes the budget constraint across all possible installation types and locations; constraint in Equation (2c) states that no more than one facility can be installed at one location. Constraint in Equation (2d) ensures that flow j can be inspected by a type- m facility only if this flow passes at least one such facility. Constraint in Equation (2e) states that each flow is inspected in one and only one scenario (including the “no inspection” scenario). Constraint in Equation (2f) is the integrality constraint on all decision variables.

The model described above is a generalization of the traditional maximum covering model, where there is one type of installation (i.e., $M = 1$), the installation cost is independent of location (i.e., $c_{im} = \text{constant}$), and $q_m = 0$ for all m (i.e., $b_{j\mathbf{k}} = h_j f_j$, $\forall \mathbf{k} \neq \mathbf{0}$; see Daskin, 1995, for a review). The model in Equations (2a)–(2f) can be further extended to accommodate realistic inspection requirements. For example, railcar flow j may have a higher inspection priority compared with others, so the inspection benefit f_j shall also be higher, and a minimum inspection frequency r_j can be imposed for railcar flow j , such that $\sum_{\forall \mathbf{k} \in \{0,1\}^M} k_m y_{j\mathbf{k}} \geq r_j$.

As described here, the model can be used to identify locations in the network that will yield the maximum benefit from inspection of different railcars, but a number of variations are possible. For example, a set of optimal solutions can be developed for railroads to guide their installation decisions under various different budget constraints. A related dual problem can find locations with the minimum installation investment possible, while achieving any specified number (or percentage) of cars in the fleet of interest that need to be

inspected in a given time period. The model can also be used to optimize the installation sequence over a period of years as budget or logistical constraints dictate, or to determine the optimal location of the next site(s), given the locations of a set of previously installed detectors. Other formulations are also possible based on various different questions of interest or constraints. Interested readers are referred to Daskin (1995) for a review of applications of some variants of the covering models.

2.2 Lagrangian relaxation

The network model in Equations (2a)–(2f) is a typical integer program and is known to be NP-hard (nondeterministic polynomial-time hard). [Correction added after online publication (Jan 14 2009): Equation reference corrected.] It could be solved using an off-the-shelf commercial solver such as CPLEXTM. However, most practical instances of this problem are likely to have a very large scale, that is, a large number of candidate locations, $|I|$, and a large number of railcar flows, $|J|$, where notation $| \cdot |$ indicates the total number of elements of a set (i.e., cardinality). In these cases, obtaining an exact solution often requires excessive computation time. Therefore, we developed an LR-based heuristic algorithm that is known to solve similar problems efficiently (Fisher, 1981; Daskin, 1995).

The basic idea of the LR method is to relax the original problem by removing certain hard constraints while adding penalty terms to the objective function for any violations to the removed constraints. Following conventional procedures, we relax the $|J| \times M$ constraints in Equation (2d) and move them to the objective function in Equation (2a) with a set of positive multipliers $\mathbf{u} = \{u_{jm} : \forall j, m\}$. This yields the following LR subproblem:

$$\begin{aligned} \text{Max}_{\mathbf{x}, \mathbf{y}} z(\mathbf{u}) &= \sum_j \sum_{\mathbf{k}} b_{j\mathbf{k}} y_{j\mathbf{k}} + \sum_j \sum_m u_{jm} \\ &\times \left(\sum_i a_{ij} x_{im} - \sum_{\mathbf{k}} k_m y_{j\mathbf{k}} \right) \\ \text{s.t.} & \text{Equations (2b), (2c), (2e), and (2f)}. \end{aligned} \quad (3)$$

The objective function in Equation (3) can be rewritten as the following:

$$\begin{aligned} z(\mathbf{u}) &= \sum_j \sum_{\mathbf{k}} \left(b_{j\mathbf{k}} - \sum_m u_{jm} k_m \right) y_{j\mathbf{k}} \\ &+ \sum_i \sum_m \left(\sum_j a_{ij} u_{jm} \right) x_{im} \end{aligned} \quad (4)$$

For any given \mathbf{u} , it can be optimized with regard to \mathbf{x} and \mathbf{y} , respectively. Regarding \mathbf{x} , the optimization of $\sum_i \sum_m (\sum_j a_{ij} u_{jm}) x_{im}$ subject to Equations (2b), (2c),

and (2f) resembles a 0–1 knapsack problem, where item (i, m) has value $(\sum_j a_{ij}u_{jm})$, cost c_{im} , and total budget B (see Martello and Toth (1990) for a review of knapsack problems). Constraint in Equation (2c) requires that at most one item (i, m) can be selected from $\{x_{im} : \forall m\}$ for each i . This problem can be solved efficiently by dynamic programming with stages $0, 1, 2, \dots, |I|$ and states $0, 1, 2, \dots, B$ (see Bertsekas (2005) for a thorough introduction of dynamic programming). Defining function $v(i, w)$ to be the maximum value achievable at stage i and under state w , they satisfy the following simple recursion:

$$v(i, w) = \begin{cases} \max \left\{ v(i-1, w), \max_m \left\{ \left(\sum_j a_{ij}u_{jm} \right) + v(i-1, w - c_{im}) \right\} \right\}, & i = 1, \dots, |I|, w = 1, \dots, B; \\ 0, & i = 0 \text{ or } w = 0 \end{cases}$$

After running the recursion for $i = 1, \dots, |I|$, $w = 1, \dots, B$, the result of $v(|I|, B)$ gives the optimal value of $\sum_i \sum_m (\sum_j a_{ij}u_{jm})x_{im}$, and the optimal solution \mathbf{x} is obtained by back-tracking.

Regarding \mathbf{y} , the optimization of $\sum_j \sum_{\mathbf{k}} \times (b_{j\mathbf{k}} - \sum_m u_{jm}k_m)y_{j\mathbf{k}}$ subject to Equations (2e) and (2f) is simple. The constraints mandate that there is only one $y_{j\mathbf{k}}$ that takes value 1 for each flow j . To maximize the objective, we simply assign $y_{j\mathbf{k}^*} = 1$ if $\mathbf{k}^* = \arg \max_{\mathbf{k}} (b_{j\mathbf{k}} - \sum_m u_{jm}k_m)$, and 0 otherwise.

The solution for optimal \mathbf{y} , for each j , requires $O(MP)$ calculations and $O(P)$ operations for finding the maximum value. The total solution for all \mathbf{y} then requires $O(MP|J|)$. The 0–1 knapsack problem is pseudopolynomial and takes $O(BM|I|)$ operations. Thus, the LR subproblem can be solved for any given \mathbf{u} in $O(MP|J| + BM|I|)$ time.

2.3 Implementation of the LR algorithm

For each given \mathbf{u} , the optimization of the LR subproblem $z(\mathbf{u})$ yields an upper bound of the optimal objective value of the original problem. It is therefore desirable to seek the smallest possible upper bound $z(\mathbf{u})$ in the hope that it will possibly equal (or be very close to) the true optimum. Mathematically, we solve another optimization problem regarding the nonnegative multipliers \mathbf{u}

$$\begin{aligned} & \text{Min}_{\mathbf{u}} z(\mathbf{u}) \\ & \text{s.t. } \mathbf{u} \in \mathbf{R}_+^{M|J|} \end{aligned}$$

We solve this problem using the conventional subgradient method (Daskin, 1995). The major steps are as follows:

Set the initial multipliers $u_{jm} = 0$ for all j, m . At each iteration n , Let \mathbf{u}^n be the current multiplier, and $t^n > 0$ be the current step size:

(1) Update the multipliers:

$$u_{jm}^{n+1} = \max \left\{ 0, u_{jm}^n - t^n \left(\sum_i a_{ij}x_{im} - \sum_{\mathbf{k}} k_m y_{j\mathbf{k}} \right) \right\}, \forall j, m$$

(2) Update step size:

$$t^{n+1} = \frac{\lambda^n (z(\mathbf{u}^n) - z^{LB})}{\sum_{j,m} \left(\sum_i a_{ij}x_{im} - \sum_{\mathbf{k}} k_m y_{j\mathbf{k}} \right)^2}$$

where λ^n is a control parameter normally starting with a value of 2. If $z(\mathbf{u})$ does not improve in a specified number of iterations, decrease λ^n slightly.

Within each iteration, the optimum solution to the relaxed LR subproblem automatically gives an upper bound of the optimal objective value of the original problem. If this solution to the LR subproblem turns out to satisfy constraint in Equation (2c), that is, be feasible to the original problem, then it is also the optimal solution to the original problem; the iteration stops. Otherwise, the iteration continues.

After each iteration, a heuristic feasible solution can also be constructed to provide (or update) a lower bound. To do this, we simply install all facilities that are determined from the LR subproblem, and then assign 1 to all $\{y_{j\mathbf{k}}\}$ according to the combination of facility types that each flow passes. This solution is feasible for the original problem and must yield a lower bound. If the lower bound equals the upper bound at any iteration, then the optimal solution is found. Otherwise, the difference between these bounds provides an optimality gap—the difference between the true optimum and the feasible solution is sure to be no larger than this gap, as the true optimal solution must be between the upper and lower bounds.

3 CASE STUDIES

An analysis of all the railcars operating on the entire U.S. railroad network, or even those on one of the major class 1 railroads, represents a huge computational effort. There are approximately 1.5 million railcars in service. One well-known national rail network model, the Princeton Transportation Network Model (PTNM), has over 50,000 nodes. The amount of data with these dimensions would pose a formidable computational challenge even if advanced data storage techniques are used

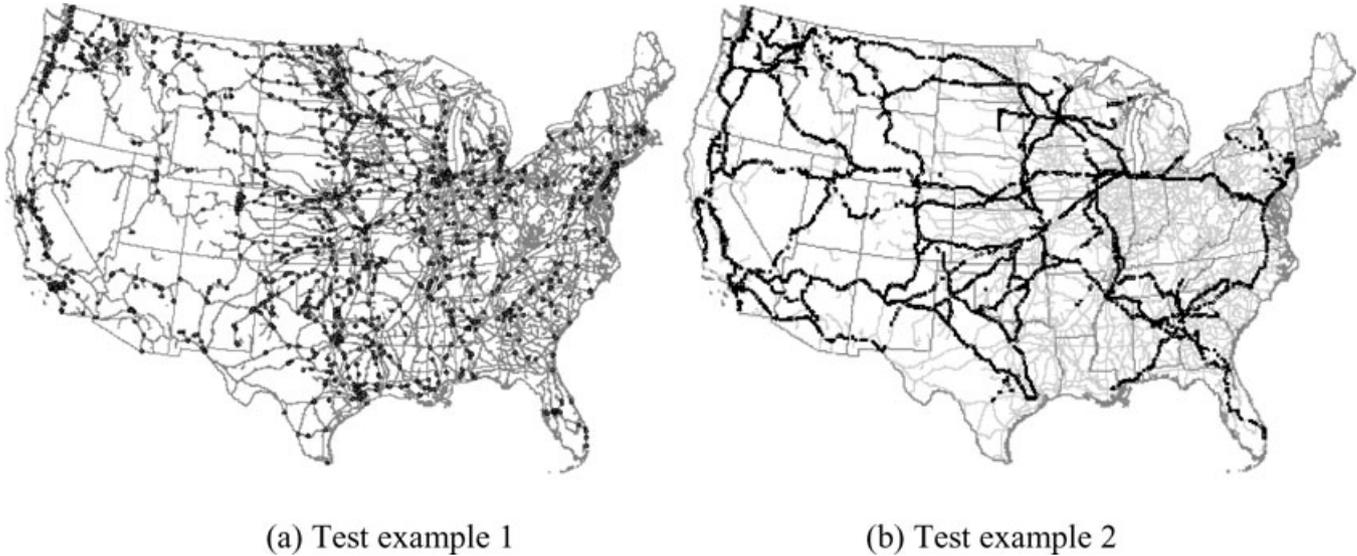


Fig. 1. The railroad network and candidate locations for the test problems.

(e.g., sparse matrix). However, there are a few techniques that can make the problem more manageable.

On many routes, the difference in the composition of traffic (i.e., the subset of railcars) varies little from node to node. In fact, on some through lines, there may be almost no change in traffic composition over hundreds of miles along a route. To reduce the dimension of the matrix, a node i can be eliminated from consideration if (1) its traffic composition is a subset of that of another node i' and (2) the installation cost $c_{im} \geq c_{i'm}$ for all m .

3.1 Testing examples with hypothetical data

3.1.1 Busy locations versus optimal locations. An empirical example was used to demonstrate the potential of the proposed model and why simply selecting the busiest locations in the network (i.e., those with the largest traffic volumes) yields suboptimal results. We selected a random sample of network locations, I , and railcar trips, J , for one of the class I railroads and flowed them over the rail network. Each year, a railcar will typically travel in many trains over multiple routes. Therefore, we identify each railcar as a distinct flow unit. The set of candidate locations I includes origin, junction, and termination locations of the railcar flows on the railroad network.

The locations where each railcar originates, passes by, and terminates can be identified using a network flow model, such as the PTNM. In PTNM, the “NET3” numbers indicate the individual nodes or locations on the network. The matrix $\mathbf{A} = [a_{i,j}]$ is formulated by determining the sample of railcars of interest and possible candidate inspection locations. In this case, we identi-

fied 8,920 individual railcars traveling around the network and 1,820 candidate locations, resulting in a corresponding matrix ($\mathbf{A}_{1820 \times 8920}$) for our analysis. Figure 1a illustrates the network (lines) and candidate locations (dots).

We first consider installing one type of wayside technology ($M = 1$) to maximize the total number of distinct railcars inspected in a year, that is, $b_j = 1$, for all j , and $q_m = 0$, for all m . In addition, we assume that $c_{im} = 1$ (million dollars) for all i, m , and the budget is $B = 10$ (million dollars). The proposed algorithm yields optimal solution within 10 CPU minutes on a 2.3 GHz personal computer (PC). At optimality, 4,951 cars (or 57% of the total railcars) would be detected by the 10 installations. It is interesting that, as discussed above, the top 10 optimal locations are not the 10 busiest nodes. In fact, only 4 out of the 10 busiest locations are part of the optimal solution. If detectors were placed at the 10 busiest locations, only 4,160 cars would be detected, which is 19% less than if the optimal set of sites were used. Although in practice, an experienced railroad expert would do better than choosing the 10 busiest locations, this simple example demonstrates the importance of optimizing detector locations at the network level. Section 3.2 also verifies that the model significantly improves the state of industry practice.

3.1.2 Expected number of correctly inspected railcars.

To test the algorithm for the problems with multiple technologies, a new sample of 4,500 network locations and 5,768 railcars was generated from the PTNM. Figure 1b illustrates the candidate locations. We now use the LR-based algorithm to solve the problem with

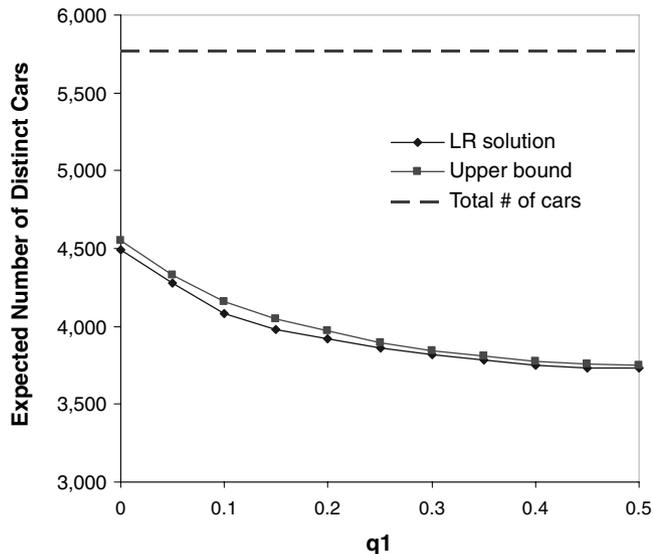


Fig. 2. Summary of LR algorithm results for varying q_1 .

multiple types of facilities, each with a different inspection performance. Consider two types of wayside technologies ($M = 2$). Suppose type 1 technology is more economical, with cost $c_{i1} = 1$ (million dollars) for all i , although it may miss detecting a defect with probability ($q_1 = 0.05$). Type 2 technology is more costly, with $c_{i2} = 2$ (million dollars) for all i , but its technical performance is fully reliable ($q_2 = 0$). We assume that $f_j = 1$, for all j , and the total budget is $B = 8$ (million dollars).

The LR algorithm yields a near-optimal feasible solution within less than 30 CPU minutes on a 2.3 GHz PC. This feasible solution has an objective value of 4,275.9 and an optimality gap of about 1.2% (upper bound 4,327.3). With our hypothetical data, eight type 1 facilities are installed, whereas no type 2 facilities are installed. This implies that the slightly inferior performance of type 1 technology is acceptable, given its substantially lower price.

Now we run some variant cases to test the influence of facility performance on the optimal solution. Suppose that the performance of the type 1 detector drops to $q_1 = 0.20$, while all other data remain unchanged. The LR algorithm yields near-optimal feasible solutions, where two type 2 facilities are now installed. The slightly inferior performance of type 1 technology makes it less appealing.

We further vary the performance of the type 1 detector from $q_1 = 0$ to $q_1 = 0.50$. The objective values (i.e., the expected number of correctly inspected railcars) of the LR solutions and the corresponding upper bounds for the true optima are illustrated in Figure 2. The dashed line on the top shows the total number of railcars in the system.

It is evident that as q_1 increases, the objective value of the feasible LR solution gradually decreases. The difference between the LR solution and the upper bound never exceeds 2%. As the true optimum is always between these two curves, we are certain that the LR solution is within a 2% optimality gap from the true optimum, for all cases.

3.2 Empirical application

We obtained empirical data from a major U.S. railroad on its network (nodes and links) and traffic (railcar shipment schedules) for 30-, 60-, and 90-day intervals.² The standard maximum covering model and LR algorithm were applied to solve for a range of 1 to 20 installation locations that maximize the number of unique railcars inspected.

The original data contain more than 10,000 candidate locations in the network, about half a million distinct railcars conducting about 2 million shipments per month. Because of the large scale, preprocessing was conducted to eliminate dominated candidate locations and merge railcar flows with the same itinerary. To eliminate dominated locations, we denote F_i as the subset of railcar flows passing location i . If $F_i \subseteq F_{i'}$ for some $i, i' \in I$, then location i is dominated by location i' and can be excluded from consideration. [Correction added after online publication (Jan 14 2009): Typographical error in subset corrected.] This would not affect the optimal solution because all the railcars that can be potentially inspected by installing at location i could have been equivalently inspected by installing at i' . To merge railcar flows, we denote L_j as the subset of all candidate locations that railcar flow j passes. If $L_j = L_{j'}$ for some $j, j' \in J$, then flow j and flow j' have exactly the same itinerary and can be merged into one railcar flow whose new volume equals $h_j + h_{j'}$. [Correction added after online publication (Jan 14 2009): Typographical errors in subsets corrected.] Also, the huge amount of data is stored in a sparse matrix format and integrated into the LR algorithm to save memory and increase processing speed. To further improve the efficiency of the LR algorithm, we temporarily store the values of the Lagrangian multipliers at convergence. These multiplier values can be used as the starting multiplier values for similar problem instances (e.g., after we slightly vary the installation budget).

A stand-alone computer program was developed for this empirical study to find and display the best set of locations that inspect the maximum number of railcar flows. The software can also determine the subset of railcars that are inspected by any given set of locations. For more information about this software, see Li and Ouyang (2007).

For the same set of network locations and railcar flows, computational time increases slightly with the number of installations. However, on a PC with a 2.3 GHz CPU, the designed LR algorithm can yield near-optimal solutions (optimality gap threshold $\approx 3\%$) in about 1 hour for all computed cases. The objective function values (i.e., the number of inspected distinct cars) are quite close for 30, 60, and 90 days of traffic. The optimality gap can be further reduced by increasing computational time, but the marginal computational effort needed increases as the gap itself gets closer to 0. In our large-scale application, if we reduce the tolerable optimality gap from 3% to 2%, the extra computational time for each problem instance is about 1 hour on average.

The railroad also provided information on its current wayside detector installations. Compared with the existing installations on this railroad's network, the solution from the proposed model (with the same number of installations) will improve the inspection benefit by a relative amount ranging from 20% to 60%.

4 DISCUSSION: POTENTIAL USES OF THE MODEL

The optimization model presented in Section 2 can be adapted to address several different types of practical problems that are of interest to the railroad industry. The following subsections discuss some of these applications, which are closely related and similar from a modeling point of view (i.e., based on a similar underlying mathematical model). However, these adaptations extend and enhance the model's utility to railroad practitioners and thus broaden its potential benefit.

4.1 Optimized wayside inspection installation by railroads

The model can be used by railroads wishing to optimize their investment in wayside inspection technology so as to maximize the utility they derive. A railroad will typically have a specified annual budget for investment in wayside detection installations. The use of the model described here would allow them to maximize the return on investment in terms of the numbers of different cars inspected. Furthermore, installation of existing technologies is likely to take a number of years, and as new technologies are developed, they will need to know the optimal deployment locations and schedules for these. The model can be used to identify the optimal sequence of facility installations over a multi-year period, given a constrained capital-spending budget, thereby helping ensure that the railroad receives

maximum benefit in each successive year of the capital plan, and thereafter, when all the installations are complete. As traffic patterns change, installation of new facilities should reflect these changes. New traffic data or projections can be used to rerun the model to check and see if existing facilities should be supplemented by new installations, or possibly moved, so that the railroad continues to maximize the benefits it derives from them.

The same approach to a railroad optimizing its installation locations can be applied at the national network level. There are four principal railroads that operate principally in the United States. Two of them have overlapping networks in the eastern one third of the nation and the other two cover the western two thirds (Figure 1). A substantial percentage of railroad freight shipments traverse more than one of these railroads between their origin and destination points. Consequently, the railcars involved in these shipments often pass wayside inspection sites on more than one railroad. North American railroads and car owners already cooperate so as to maximize the value of wayside inspection data collected from sites throughout the network in the form of the AAR's Integrated Railway Remote Information Service (InteRRIS) program (Irani et al., 2003a; Hawthorne et al., 2005). Data are collected and analyzed for individual railcars from wayside inspection facilities throughout the network and analyzed at a centralized location. The reports are then communicated to various individual railroads and car owners that subscribe to the service. The consolidation and centralization of the data processing enables trending and potentially earlier warning to railcar owners for data for individual railcars than would otherwise be possible.

In the same manner that an individual railroad might want to optimize the placement of inspection facilities, the railroads collectively may wish to do so as well. It is almost certainly the case that optimizing at the national level will be more cost-effective overall than each railroad only considering locations on its own property. Consider two high-density mainlines on two different roads that interchange traffic at a common gateway. Optimizing locations individually, they might each locate an inspection facility on their respective side of the interchange point. If they are interchanging much of this traffic, then one installation used by both might be nearly as beneficial. If a single installation provides nearly as much benefit as two, then the optimization model can be used to identify another location where an installation would provide greater marginal benefit to the industry as a whole. Avoiding this kind of duplication frees up capital for installations at other locations or other uses. An interesting question is how different the set of locations might be when optimizing at the national level compared with the collective set of sites, if

each railroad optimized based on its own traffic and network alone, and if there are substantial potential savings from such a collective approach. This is a literal example of “global” versus “local” optimization.

4.2 Optimized inspection of certain railcar or cargo types

A more specialized application of the model’s potential use involves particular types of inspection technologies that may be particularly important for certain types of cars or traffic. For example, certain railcar types are particularly susceptible to problems with the ability of their trucks to steer well in curves. These cars can cause derailments in certain curve alignments, even though their performance is within the AAR specifications. A detector placed ahead of such curves might identify a problem, and the train could be stopped. However, although this could prevent a derailment, it would delay the train, and all the cars in it, thereby disrupting service. A better solution would be to ensure that the performance of the components affecting steering are being monitored, and repairs or maintenance take place well in advance of any potential to cause a problem at a time and location that could be better planned and managed. This is part of the underlying philosophy of InteRRIS, but it also applies to the strategic placement of wayside detectors. As the car types that are prone to these problems are known, along with their typical routing patterns, their identity can be a parameter in the model used to select locations for the particular type of wayside installation that monitors the type of problem they are susceptible to.

A related example is cars transporting cargoes whose on-time delivery is particularly important. They may be transporting high-value commodities or time-sensitive hazardous materials where the delay due to repair of a failing component would be particularly costly, troublesome to the customer, or pose risk. Or, it might be a car transporting a hazardous material where an accident has the potential to cause severe damage or injury. In these cases, there may be an additional premium for detecting and correcting components in time for planned maintenance to occur before they cause a serious delay or accident. Again, the model could be used to optimize the placement of relevant types of detector, taking into account the traffic routing and density patterns of these particular railcars and commodities.

Delays of such shipments, particularly due to accidents, will often be affected by problems on other cars. An alternative version of the model might be constructed in which the objective function is to minimize the likelihood of accident occurrence or minimization of delays in general, and then compare the solutions with one developed based on the objective function de-

scribed in this article. If they were different, then it might suggest an alternative placement strategy, and if they were similar, it would enhance confidence in the value of the solutions developed by the model using the objective function, as described here.

4.3 Assessing the performance of inspection facilities

The model may also have utility in providing railroads with objective performance assessments of their overall inspection activity. Railroads have already deployed wayside inspection equipment at various points around their systems based on a variety of criteria. These are providing useful information, and additional installations are underway. However, it is difficult for railroads to know how the benefit they are deriving compares with some objective expectation of the maximum they might be able to achieve from their investment in this technology. The model can be used to provide an objective, quantitative metric of the potential performance that they can compare with the performance of their current installations. As mentioned above, if they are unsatisfied with the results from such a comparison, they can use the model to guide further new installations or redeployment of existing equipment in a more efficient manner.

5 CONCLUSIONS

Wayside inspection technology is widely used and is critically important to railroads to maintain safe and efficient operations. There is rapid development and deployment of new technologies that have greater capabilities than earlier generations of detectors (Barke and Chiu, 2005; Lagnebäck, 2007). These new technologies are more capable of identifying performance metrics that are predictive of failure and are useful for longer term trending analyses of individual railcar performance. However, these new technologies are considerably more costly, and railroads are often capital-constrained, so it is important that railroads optimize both the locations and the sequencing of these installations so as to maximize the utility they derive.

In this article, we present a network design model to optimize the installation placement of wayside technologies. The formulation allows consideration of options, such as locations, technology types, and costs and benefits, to be specified. We also develop a heuristic method based on the LR. The numerical examples show that it yields near-optimal solutions in reasonable computational time. On the basis of the model, a stand-alone computer program was developed that solves this problem for any large-scale railroad network.

In the future, the stand-alone software will be further enhanced so that railroads can use it directly for various purposes. Metaheuristics that have the potential to further improve the solution will be considered and included in the software. On the modeling side, several assumptions are currently made in the article (e.g., on technology performance probability and independence). Future work shall consider possible relaxation of these assumptions. Also, future railcar flows (e.g., over a longer time horizon) may be unknown, and stochastic factors can be incorporated into the model. Overall, the research results should help railroads better allocate limited resources and coordinate with each other to maximize the utility they derive from installations of new, advanced-technology, wayside inspection systems. Possible application of the proposed sensor location model to other transportation networks (e.g., highway network and intermodal networks) for other objectives (such as traffic monitoring and travel time estimation) should be explored in future research.

As a final remark, the predictive detecting technologies look for irregularities in performance of a component that suggest that some type of maintenance may be needed in the future. Related to the development of these technologies is the ability to integrate information from many detector sites as they inspect the same railcar at various points as it travels around the rail network. Systems such as InteRRIS consolidate and analyze this type of data and identify trends that may suggest a developing problem (Irani et al., 2003a, 2003b). It will be interesting to explore the possibility of extending the models developed in this article to address this issue.

ACKNOWLEDGMENTS

This research was supported in part by research grants to the first author from the Association of American Railroads (AAR) Technology Scanning Program and CSX Transportation, Inc. We thank Alain L. Kornhauser and ALK Technologies, Inc., for providing the PTNM software, and Dharma Acharya, Natasha Zurikat, and CSX Transportation, Inc., for data support.

NOTES

1. The definition of facility types can be general; that is, it can be a certain type of wayside technology or a combination of multiple technologies.
2. Because of data confidentiality agreements, we cannot disclose details of the input data and solution output.

REFERENCES

- Association of American Railroads. (2006), *Railroad Facts*, Association of American Railroads, Washington, DC.
- Austin, K. B. (1949), Hotbox signal for railway trains, U.S. Patent 2486546.
- Barke, D. & Chiu, W. K. (2005), Structural health monitoring in the railway industry: a review, in *Structural Health Monitoring*, vol. 4, pp. 81–93.
- Bertsekas, D. P. (2005), *Dynamic Programming and Optimal Control*, vol. 1, 3rd edn., Athena Scientific, Nashua, NH.
- Blevins, W., Morgan, R. & Pinney, C. (2003), Integrated wayside systems to improve safety and efficiency, in *Proceedings IHHA 2003 Specialist Technical Session – Implementation of Heavy Haul Technology for Network Efficiency*, May 5–9, 2003, Dallas, TX.
- Burpee, C. M. ed. (1945), *Railway Engineering and Maintenance Cyclopedia*, 6th edn., Simmons-Boardman Publishing Company, Chicago, IL.
- CN. (2008), Safety: technology – hot bearing detectors (hotbox), available at: <http://www.cn.ca/about/safety/safety/equipment/en_HBD.shtml>, accessed January 20, 2008.
- Daskin, M. S. (1995), *Network and Discrete Location: Models, Algorithms and Applications*, Wiley, New York, NY.
- Edwards, J. R., Barkan, C. P. L., Hart, J. M., Todorovic, S. & Ahuja, N. (2006), Improving the efficiency and effectiveness of railcar safety appliance inspection using machine vision technology. JRC2006-94052, in *Proceedings of the ASME Joint Rail Conference*, Atlanta, GA, April 2006.
- English, G. (1996), *TP 12691-E, Hot Bearing Detector Study Railway Safety Directorate*, Transport Canada, Ottawa, Ontario.
- Fisher, M. L. (1981), The Lagrangian relaxation method for solving integer programming problems, *Management Science*, **27**, 1–18.
- Gallagher, C. A. & Pelino, W. M. (1959), Hot-box detector, U.S. Patent 2880309.
- Hawthorne, K., Wiley, R. B. & Guins, T. (2005), Wayside monitoring with ATSI aids wagon maintenance, *Railway Gazette International*.
- Irani, F., Tournay, H., Tunna, J. & Cummings, S. (2003a), Need stress relief? By bringing wayside defect-detector data together, TTCI's inteRRIS™ database will be able to provide the North American rail industry with preventive and predictive maintenance of its freight car fleet and lower railroad stress-state, *Railway Age*, December 2003.
- Irani, F. D., Wiley, R. B. & Guins, T. S. (2003b), Network efficiency gains using a detector-based information management system, in *Proceedings IHHA 2003 Specialist Technical Session – Implementation of Heavy Haul Technology for Network Efficiency*, May 5–9, 2003, Dallas, TX.
- Lagnebäck, R. (2007), *Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles*. Licentiate Thesis, Luleå University of Technology, Department of Civil, Mining and Environmental Engineering Division of Operation and Maintenance Engineering, Luleå, Sweden.
- Lai, Y. C., Ahuja, N., Barkan, C. P. L., Drapa, J., Hart, J. M., Jawahar, C. V., Kumar, A., Milhon, L., Narayanan, P. J. & Stehly, M. (2007), Machine vision analysis of the energy efficiency of intermodal freight trains, *Journal of Rail and Rapid Transit*, **221**, 353–64.

- Li, X. & Ouyang, Y. (2007), *Railroad Wayside Detector Location Solver (RWDLS) Software: User's Guide*, University of Illinois at Urbana-Champaign, Urbana, IL.
- Lundgren, J. (2007), Enhancing railway safety and productivity with automated inspection technologies: the art and science of watching and listening to your trains, in *Proceedings: International Heavy Haul Conference – Specialist Technical Session: High Tech in Heavy Haul*, IHHA, Inc., Kiruna, Sweden; Virginia Beach, VA, 507–17.
- Lundgren, J. & Killian, C. (2005), Advanced rail vehicle inspections system, in *Proceedings 8th International Heavy Haul Conference IHHA*, IHHA, Inc., Rio De Janeiro, Brazil; Virginia Beach, VA, 807–14.
- Martello, S. & Toth, P. (1990), *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley & Sons, New York, NY.
- Post, W. M. (1936), Protective system for railways, U.S. Patent 2063336.
- Post, W. M. (1937), Protective system for railways, U.S. Patent 2095616.
- Resor, R. R. & Zaremski, A. M. (2004), Factors determining the economics of wayside defect detectors, in *Proceedings of the 83rd Annual Meeting of the Transportation Research Board* (CD), Washington, DC.
- Steets, P. G. & Tse, Y. H. (1998), Conrail's integrated automated wayside inspection, in *Proceedings of the IEEE/ASME Joint Railroad Conference 1998*, IEEE, Piscataway, NJ, 113–25.
- Tournay, H. M. & Cummings, S. (2005), Monitoring the performance of railroad cars by means of wayside detectors in support of predictive maintenance, in *Proceedings of the 8th International Heavy Haul Conference IHHA*, 14–16 June 2005, Rio De Janeiro, Brazil, 501–10.
- Tournay, H., Wolgram, T., Cummings, S. & Chapman, S. (2006), *Interpreting Truck Performance Detector Data to Establish Car and Truck Condition, R-977*, Transportation Technology Center, Inc.; Association of American Railroads, Pueblo, CO.
- Wolf, G. P. & Peterson, F. J. (1998), Application and validation of the automated truck performance measurement system, in *Proceedings of the IEEE/ASME Joint Railroad Conference 1998*, IEEE, Piscataway, NJ, 127–34.