A Mixed Integer Programming Approach to the Locomotive Assignment Problem

Francesco Piu†

Visiting PhD at the TRANSP-OR laboratory,
École Polytechnique Fédérale de Lausanne, francesco.piu@epfl.ch
†Department of Quantitative Methods, University of Brescia,
C. da S. Chiara 50 - Brescia, Italy

Abstract

Every year large railroad companies invest billions of dollars acquiring, managing and fueling locomotives, therefore even small percentage improvements toward a better efficiency in the use of locomotives, can lead to significant economic savings. The locomotive assignment problem (LAP) is solved assigning a fleet of locomotives to a network of trains optimizing one or more crucial objectives (costs, profit, fleet size, level of service, . . .) and satisfying a rich set of technical and economic constraints. Starting from a deterministic train scheduling and focusing on the planning version of the problem, Ahuja et al. propose to model the LAP as a Mixed Integer Programming (MIP) problem and to solve it as a multicommodity flow problem with side constraints (the number of locomotives of each type is limited) on a space-time network. This model, is characterized by about 197000 integer variables and 67000 constraints, consequently the problem has been solved combining several heuristic solution steps.
This work is motivated by the development of a new model able to deal with real-life aspects of the planning LAP not directly included in the considered MIP model: locomotive fueling, locomotive maintenance, uncertain schedule.

Keywords: freight train, heuristic, locomotive assignment, locomotive scheduling, mixed integer programming, multicommodity flow problem, space-time network

May 1, 2011

Introduction

The strong competition among railroads and the growing role of the private sector, specially in Europe where many national railroad proceed toward a privatization, imply that railroads are paying more and more attention on operating cost, punctuality and performance, which affects customers' satisfaction.
The U.S. freight transportation system is one of the best example of the effects of the competition among transport companies. The whole system (highways, waterways, airways, and railways) offers the best service and rates in the world and the freight rail element of this system is critical to the competitiveness of many industries and the economies of many states (Grenzeback et al. [2008]). America's freight railroads span 140,000 miles and form the most efficient and cost-effective freight rail system in the world (Thompson [2007]). Historically, U.S. and Canada offer the richest set of railways companies dealing with the highest competition rate in the world. Many of these companies invested and invest in the creation of tools like simulation and optimization models to help their decisions process. So, not surprisingly, many optimization models for the locomotive assignment problem (LAP) have been developed by North American research centers and focus on real problems faced by U.S. and Canadian companies.
In the last decades however, an increasing interest in optimization models for the LAP emerged among, for instance, European, Australian, Indian and Brazilian railway companies.

The rest of the paper is organized as follows: section 1 introduces a short historical perspective of the role of optimization models in railway's scheduling problems, section 2 presents the locomotive assignment problem in its different application fields (freight trains, passenger trains, switch engines, industrial in-plant railroads), section 3 introduces the literature reviewed, section 4 describes the mathematical model for the locomotive assignment problem adopted as reference (state of the art), section 5 summarizes the original methodological contributions of this paper. Discussion and future research are outlined in the last section.

1. Tonnage-based and scheduled-based approaches: the role of optimization models

The increased computational power in the last decade allows the tractability of more complex models and bigger instances. Consequently, the unavoidable complexity and size of real-life problems can be captured and managed more efficaciously leading to the creation of effective decision-support tools for realistic applications. But this element tells us only a part of the story.

The increasing interest in optimization models, cannot be completely explained by the increasing computational power and modeling ability. In the last two decades, passenger and freight movement over the transportation system have increased significantly in both advanced (like U.S., Europe, Japan) and emerging (like Cina, India, Brazil) countries.

The U.S. rail freight transportation system represents a significant example: the ton-miles of rail freight (one ton-mile represents one ton of freight carried over one-mile counts) moved over the national rail system have doubled since 1980, and the density of train traffic measured in ton-miles per mile of track has tripled since 1980 (Grenzeback et al. [2008]). Despite the fact that the rail's share of the total freight transportation market is moderate (14 percent of total tons carried, 25 percent of total ton-miles) and that the rail's market share is also declining, the current demand for rail freight transportation is pressing the capacity of the rail system (Grenzeback et al. [2008]).

The need to reduce greenhouse gas emissions (like CO2) will probably increase even more this demand because the freight rail service is very fuel-efficient and generates less air pollution per ton-mile than trucking (Grenzeback et al. [2008]).

Given the demand for freight transportation, usually expressed in terms of weight (tonnage), the railroad establishes a policy for the routing of trains. A possible policy is to running trains only when they have enough freight. This policy has been traditionally practiced by North American railways and is named tonnage-based dispatching. In a tonnage-based approach in dispatching trains, the company holds all trains until they have enough tonnage, so a train may be scheduled every day, but it may be delayed or canceled, depending on the achieved tonnage.

The idea underlying the tonnage-based approach is simple: to minimize the total number of operating trains by maximizing the train size in order (theoretically) to minimize the crew costs and maximize the track capacity. However, from a practical point of view, this approach presents some limitations and shortcomings:

(i) Railroad tracks used for loading/unloading, sorting, or storing railroad cars and/or locomotives (i.e. yards) cannot optimize their operations relying on a repetitive schedule, and they may require more railcars and more storage capacity to deal with the traffic variability.

(ii) It may be an increased demand for crew and locomotive resources.

(iii) Operating costs may increase due to an increased idling cost and relocation cost of equipments and crews and also due to a reduced utilization of locomotives and railcars.

Moreover, the tonnage-based approach implies an unreliable and poor service for the customers, so the tonnage-based approach was, and remains, a good strategy for bulk goods like coal, but it has proven to be a poor strategy for most other goods.
Although the tonnage-based approach is still common in North America, it is rarely used in the European context where freight trains typically operate according to schedules (like passenger trains): this is the schedule-based approach. In the schedule-based approach trains run as scheduled, even when the train has not achieved a sufficient tonnage.

The management of a schedule-based strategy implies that the schedule should be adapted depending on the forecast of the traffic and requires advanced computers and operations research software to conduct deep analyses of different alternatives in short times. As reported in Ireland et al. [2004], the scheduled-based strategies recently gained favor in U.S. and Canada where several railway companies are adopting this more disciplined approach to obtain cost-effective and customer-effective operating plans.
The discussed increase in customer demand for freight rail transport and the disposal of advanced computers and operations research software (not available until recently) push several North American railway companies to change the paradigm of their operations passing to a schedule-based strategy.

In fact running more frequent trains with scheduled transit-time may require more simultaneous operations of assembling and dispatching, so more trains are running simultaneously leading to more congested rail tracks and yards. More locomotives are required and more empty railcars need to move faster to be ready for loading (Ahuja et al. [2005a]).

At the beginning of the new century Canadian Pacific Railway (CPR) obtains more than 500 million (Canadian $) of annual operating costs savings. These savings are because of the ability to better execute the plan through daily repetition and to better manage crews and equipment (faster railcar velocity, improved locomotive utilization, reduced train starts). In addition to cost savings, running on a schedule has allowed CPR to recapture traffic from the trucks. The new schedule-based approach has allowed CPR to think and act like truckers (Cambridge Systematics Inc. [2005]).
The success of the new Operations Research tools used by CPR has (surprisingly) overturned the old paradigm that tonnage-based plans are more efficient.
In the last years all North American Class I railroads follow the example of CPR. NS and CN, switch most of their services to run on a scheduled operating plan (also CSX Transportation, and FEC have all adopted the scheduled railroading philosophy) (Cambridge Systematics Inc. [2005]).
Supporting the historical role of simulation tools, optimization models are gaining more and more importance in solving large size complex scheduling problems that characterize the schedule-based approach in real-life applications.

2. Problem description

The LAP is one of the most important problems in railroad scheduling because it involves very expensive assets and huge numbers. Every year, large railroad companies invest billions of dollars acquiring, managing and fueling locomotives. Every day they assign thousands of locomotives to thousands of trains. Due to the size of real-life problems, even a small percentage improvement toward a better efficiency in the use of locomotives, can lead to significant economic savings.
For this reason it is crucial to find a satisfactory solution to the LAP. Due to its importance, many researches focused on this scheduling problem. Unfortunately, the Locomotive Assignment Problem (LAP) is a very complex discrete optimization problem that has not been solved in a completely satisfactory way. Indeed the LAP is so complex, from a mathematical modeling and (even more) computational point of view, because of its detail richness and its size in real-life applications.

The LAP is solved assigning a fleet of locomotives to a network of trains usually minimizing the total operational cost and satisfying a rich set of constraints (both technical and economic).
Locomotive scheduling may be studied at three levels: planning (or strategic), tactical and operational, in accordance with the length of the respective planning horizon and the temporal impact of the decision. Roughly speaking, the three notions identify the planning activities in the long, mid and short term, in that order. At the strategic level only the number of locomotives and their type matter, the specific ID of each locomotive is not considered and locomotives of the same type are completely equivalent.
In the planning version of the LAP, for each train we determine the type and the number of locomotives assigned to that train. Usually in the planning LAP the train schedule is given and cannot change (no delays or disruption are considered).
On the contrary the dynamic (tactical and operational) LAP, introduce many aspects not considered in the planning version. This is necessary because we deal with specific locomotives and not just with locomotive types. More precisely, we have to assign locomotive ID codes (unique for each specific locomotive) to trains. This means that we have to solve a locomotive routing problem while honoring the constraints of the scheduling phase and new operational constraints (like fueling constraints, maintenance constraints, ...).

Furthermore the train schedule may be affected by delays and disruptions events.

The demand for a particular train is expressed in terms of tonnage and horsepower (HP) and may be satisfied choosing a feasible and suitable consist from a set of locomotives of different characteristics.
Very often, this suitable consist does not match exactly the power desired for the train and provide more power than needed. A careful choice of consists can minimize wasted power but in general cannot eliminate it. It is important to note that the wasted power is the consequence of an integral number of locomotives and disappear when a relaxed problem is solved for a fractional number of locomotives. This fact produces an integrality gap that makes the solution of this problem more complicated than that for aircraft scheduling, vehicle scheduling, vehicle routing or crew scheduling. In the solution proposed in Ziarati et al. [1997] the author faces integrality gaps generally well above 5%. This level of integrality gap is observed rarely in vehicle routing and crew scheduling contexts. Ziarati et al. report that LAP usually exhibit tight integrality gaps (the solution of the linearly relaxed problem is strongly fractional) because the train covering constraints are writtren in terms of tonnage and horsepower rather than in terms of number of locomotives (another reason reported by Ziarati et al. was provided by the theoretical results presented in Bramel and Simchi-Levi [1996]).
Moreover, Ziarati et al. stress that choosing the consist that minimizes wasted power may lead to locomotive unbalances in stations that increase locomotive non-productive movement (deadhead and light traveling). In fact, non-productive movement may increase in order to solve this locomotive unbalances i.e. locomotives arriving in a particular station may not be the same departing from that station. Then we have a trade-off between minimizing wasted power and minimizing non-productive locomotive movements and part of the problem in selecting consist is to find an optimal trade-off between these two concurrent requests.

The minimization of operational cost imposes the consideration of other crucial aspects involved in the consist selection problem. The reduction of costs is primarily pursued minimizing the number of used locomotives and so minimizing the non-productive time spent by equipment, crews, technicians and so on. It is then important to promote the use of train to train connections avoiding consist-busting operations.

A consist-busting operation is characterized by very labor, cost, and time intensive activities. Additional crews are needed to transfer locomotives between the rail yard and the shop facility. Here, initial consists are busted and reassembled to form other consist, different from the initial one. Transferring locomotives across a station introduces incremental moves for the yard masters to worry about. Car switching activity must be paused and train movements must be coordinated between different department (transportation and mechanical) as opposed to train to train connections which are entirely supervised by transportation. As a consequence, consist-busting requires between two to six additional hours per locomotive within the station (Vaidyanathan et al. [2008a]).

2.1. Freight and passenger railway transportation

Passenger and freight trains have different characteristics, passenger trains always run according to a fixed schedule while freight trains may operate without schedules and simply depart when they have accumulated a sufficient tonnage. Passenger trains are more time sensitive and so they have the priority where they share the same rail network with freight trains (a common occurrence in U.S., Canada, Europe, Australia and in many developing countries).

Typically passenger trains are lighter than freight ones since they use a small number of cars coupled with one or two locomotives while freight trains generally contain a large number of cars coupled with several engines. For passenger trains the maximum gross weight is known in advance with a small uncertainty while the weight of freight can change unexpectedly for both scheduled and not scheduled trains.

There are significant differences in complexity and modeling of the planning LAP in the passenger and freight frameworks. Very often a single locomotive is sufficient to pull a passenger train (so the load of the train no longer features in the formulation) and when more than one locomotive is needed the consist is usually constituted by no more than two locomotives of the same type. According to Noble et al. [2001], in the first case the problem is modeled assuming several classes of locomotives but a single pulling locomotive (multi-class single-locomotive problem), in the second case the train is pulled by a multi-locomotive consist (multi-class multi-locomotive problem). In both cases the reduced size of passenger trains and consist make the problem more tractable with respect to the freight version, so it is possible

to assign simultaneously both locomotives and cars to the passenger trains (Cordeau et al. [2000], Cordeau et al. [2001], Lingaya et al. [2002]), while for freight trains this two assignments are managed separately.
As reported in Cordeau et al. [1998], few works focusing on the management of passenger railway locomotives can be found in the operations research literature. Ramani [1981] focuses on the problem faced by Indian Railways, Cordeau et al. [2000], Cordeau et al. [2001] and Lingaya et al. [2002] treat the problem of simultaneous locomotive and car assignment at VIA Rail Canada, Illés et al. [2005] and Illés et al. [2006] treat the locomotive assignment at Magyar Államvasutak (MÁV) (the Hungarian State Railway Company), Maróti and Kroon [2005] study the maintenance routing of trains at NS Reizigers (the main Dutch operator of passenger trains), Paoletti and Cappelletti [2007] present a decision support system developed by the Models and the Decisional Systems Department of Trenitalia (the main Italian operator of passenger trains) to aid the locomotive fleet planning.
Resuming, we have the following table

Table 1: Passenger railway locomotive management

Authors	Railway company
Ramani [1981]	Indian Railways
Cordeau et al. [2000]	VIA Rail Canada
Cordeau et al. [2001]	VIA Rail Canada
Lingaya et al. [2002]	VIA Rail Canada
Illés et al. [2006]	Magyar Államvasutak
Illés et al. [2005]	Magyar Államvasutak
Paoletti and Cappelletti [2007]	Trenitalia

More researches focus on the more complex freight railway engine assignment. Some of the reasons that make the planning LAP for freight trains more complex are:

- The number of active locomotives is often two or three times the one required in passenger trains (consist may be constituted by 4 or more pulling locomotives, depending on the maximum number of axles permitted on a single consist)

- The number of active and passive locomotives attached to freight trains can be many times the number of locomotives attached to passenger trains (for instance, CSX imposes the maximum number of locomotives on a single train equal to 12)

- There are many different types of trains, belonging to (three) different classes (intermodal, auto and merchandize), that require very different consists, so it is more difficult to reduce the size of such a heterogeneous set of consist

- There are much more train to train connections possibilities to be considered, much more constraints (like locomotive versus train compatibility constraints) and complications like cost of coupling and uncoupling consists (consist busting)

The following table reports a (non exhaustive) list of researches inspired by real problems

Table 2: Freight railway locomotive management

Authors	Railway company
Florian et al. [1976]	Canadian National
Ziarati et al. [1997]	Canadian National
Ziarati et al. [1999]	Canadian National
Ziarati et al. [2005]	Canadian National
Ireland et al. [2004]	Canadian Pacific Railway
Ahuja et al. [2005a]	CSX Transportation
Ahuja et al. [2005b]	CSX Transportation
Ahuja et al. [2006]	CSX Transportation
Vaidyanathan et al. [2008a]	CSX Transportation
Vaidyanathan et al. [2008b]	CSX Transportation
Powell and Bouzaiene-Ayari [2007]	Norfolk Southern
Brannlund et al. [1998]	Banverket Swedish National Railway
Scholz [2000]	Stateiis Järiivägar Swedish State Railways
Noble et al. [2001]	Public Transport Corporation
Baceler and Garcia [2006]	Companhia Vale do Rio Doce
Fügenschuh et al. [2006]	Deutsche Bahn AG
Fugenschuh et al. [2008]	Deutsche Bahn AG

2.2. *Yard switching and in-plant railroad LAP*

Railroad yards are a complex series of railroad tracks for storing, sorting, or loading/unloading railroad cars and/or locomotives and constitute crucial components of a railroad network. They are the points of origin and destination of shipments and freight movements. In a yard, inbound trains are disassembled, unloaded and inspected. After that (when needed) cars and locomotives are sent to cleaning and maintenance cfacilities (shops). Finally they are loaded and reassembled forming new outbound trains.
As reported in Sabino et al. [2010], yard activities are an important part of freight transportation operations since the delays associated with these activities represent a large portion of the transit time for rail freight. Yard locomotives are often called switch engines, they move cars and locomotives within the railroad yard. The objective is to minimize the costs of the switch operations and the solution of the LAP helps to optimize the fleet size of the switch engines that greatly affect this costs, see Sabino et al. [2010] for more details.

Lübbecke and Zimmermann [2003] report the LAP faced by another particular railroad sector. Large industrial plants in the automobile, chemical, and steel industry require to transport freight between production, storage, or shipping terminals that are often widely spread. In order to preserve a timely production process it may be useful to have a private railroad system which manage this tasks (often a subsidiary and a distinct legal entity). An industrial in-plant railroad has to be managed minimizing operational cost and the LAP has to be solved efficiently. There are very few studies dedicated to this particular version of the LAP, one of the first is Charnes and Miller [1956], for a more recent research see Lübbecke and Zimmermann [2003] in which real application at Verkehrsbetriebe Peine-Salzgitter GmbH and EKO Transportgesellschaft GmbH are presented.

3. Problem types

The locomotive assignment problem may be classified into a several kind of categories depending on the classification parameter that is considered. For instance, problems can be classified looking at the *planning level* and so the problem may be a strategic, tactical or operational locomotive scheduling. Another possibility is to classify problem by the objective pursued by the modeler, for instance: minimize operating costs (maximize profits), minimize deadheading times, minimize fleet size, and so on.

From a modeling perspective a first important classification can be obtained considering as classification parameter, the maximum number of pulling locomotive a train may require. If each train in the problem needs a single pulling locomotive then the problem is modeled by a *single locomotive model* (or single engine model). If some trains require more than one pulling locomotive then the problem is modeled by a *multiple locomotive model* (or multiple engine model).

3.1. Single locomotive models

Ceteris paribus, the problems in the single locomotive category are clearly easiest to solve. It is natural to proceed further in the classification considering how many locomotive types the model requires. If the problem is modeled assuming only one type of locomotive, then it becomes similar to the single depot (bus) vehicle scheduling problem (SDVSP), if many locomotive types are required, then the problem is similar to the multiple depot (bus) vehicle scheduling problem (MDVSP) (Forbes et al. [1991]). The first version (SDVSP) may be modeled as as a minimum cost flow problem whose solution is achievable for very large scale instances as remarked in Ziarati et al. [1997] since it can be solved efficiently by polynomial or pseudopolynomial algorithms, for instance by the so called Hungarian Method (Fügenschuh et al. [2006], see Ahuja et al. [1993] for details). As the former, the vehicle scheduling problem with multiple depot has been widely studied.

Focusing on the railway area, Booler [1980] considers a one day cyclic train schedule with possibly variable trains departure times and propose a model based on multi-commodity flows. The objective is to find a minimum cost set of locomotive schedules to pull a given set of trains. Booler proposes a heuristic method based on a linear programming model. Booler tests the method on small instances (10 to 50 trains) and Wright [1989] point out that this approach does not produce good solutions for larger (more realistic) instances (100 to 500 trains).

Wright seems the first author able to find a valid solution for a large-scale instances of this problem. He consider a cyclical one day train schedule and obtain the solution through a heuristic procedure. Wright test the procedure

on several instances (25 to 200 trains) however he does not take into account the existing constraints for the fleet size, for this reason Wright does not recommend the use of this procedure for real life applications.
Forbes et al. [1991], inspired by the work of Wright, obtain an exact solution for the locomotive scheduling problem. They translate to the locomotive scheduling problem, a solution procedure they developed for the MDVSP in a previous work: the model is based on an integer linear program equivalent to a multi-commodity flow formulation where each commodity represents a locomotive type. This method represent a significant improvement over the method proposed by Wright, also because Forbes et al. are able to take into account the fleet size constraints, not included in the model of Wright.

More recently Fügenschuh et al. [2006] follow a path similar to the one adopted in Forbes et al. [1991]. They start from their experience on multi-depot multi-vehicle-type bus scheduling problems and extend their and other authors solution methodologies to the locomotive scheduling problem. As Forbes et al., Fügenschuh et al. point out the extra difficulties of locomotive scheduling problems due to the several new aspects that have to be taken into account: cyclic departures of the trains, time windows on starting/arrival times, transfer of wagons between trains. The model is formulated as a linear integer programming problem, in two different versions: with fixed and with flexible starting/arrival times.
The fixed starting time version of the problem is called *capacitated cyclic vehicle scheduling problem* (CVSP) due to the cyclic character of the locomotives schedules. The capacity of the vehicles represents an upper bound for the availability of the different locomotive types. The flexible starting time version is called *cyclic vehicle scheduling problem with time windows* (CVSPTW). The CVSPTW is further specialized in two sub-versions, the first consider constant traveling times while in the second the driving time of the trains is not constant but depend on the total network load. This take into account the fact that often freight and passenger trains share (most of) the same tracks and so at daytime a freight transport may wait for the passenger transport and then the average traveling speed may be much lower than at nighttime. Their work aims to improve the simulation tool used by the Deutsche Bahn AG, the largest German railway company. Their model is based on a multi-commodity min-cost flow formulation and is solved as a linear integer programming problem.
The CVSP and the CVSPTW problems are formulated as integer program-

ming problems and commercial IP solvers (ILOG Cplex 10) are used to compute feasible/optimal solutions.
Fügenschuh et al. are able to solve instances of the CVSP up to 1537 trips and 4 locomotive classes while for the CVSPTW they consider up to 120 trips and 4 locomotive classes and time windows lenght ranging from ± 10 to ± 120 minutes intervals around the pre-scheduled starting time.

As Charnes and Miller [1956] before, Lübbecke and Zimmermann [2003] treat the in-plant railroad engine scheduling and routing problem, a subject that has not been extensively discussed in the operations research literature. They describe the mathematical and algorithmic solutions proposed to in-plant railroads as decision support tools for scheduling and routing problems. The minimization of the total deadheading and waiting time is considered as an example of practically relevant objective function. The problem is related to the multiple-vehicle pickup and delivery problem and two formulations of the problem are considered: a mixed integer and a set partitioning programs. The linear programming relaxation of the set partition model is solved by column generation. Computational experiments on both artificial and real-life data from three different German plants (VPS, EKO and SOL).

3.2. Multiple locomotive models

The most complete version of the LAP occurs when consists, (instead of single locomotives) are linked to trains and there is more than one locomotive type, so a single train may be linked with several locomotives of different types. This is the LAP with heterogeneous consists.

Florian et al. [1976] analyzed a freight train problem for Canadian National Railways (CN) and were among the first to deal with this version of the problem. They formulated the problem as an integer program based on a multi-commodity network flow formulation. The objective is to minimize the capital investment and the maintenance costs over a long planning horizon selecting an optimal number of (mixed) engine types that satisfy the motive power requirements of each train. The power requirement constraints are determined according to train weight, train length (number of cars) and geography (rule grade of traveled track).
They propose a solution based on a Benders decomposition method and conduct computational experiments using the weekly train schedule for the

Atlantic region of the CN. Their implementation does not converge rapidly so the problem could not be solved to optimality and the size of the optimality gap was considered acceptable for medium-sized problems but not for large ones. It should be noticed that the limited computational power imposed to stop the algorithm after less than 30 iterations, different convergence result could be probably obtained with the present computers.

Ziarati et al. [1997] extended the formulation proposed in Florian et al. [1976] to include many of the operational constraints encountered at CN (e.g. active and passive deadheading, scheduling of the maintenance intervals of the locomotives, noncyclic trains schedules with fixed start and ending times). Ziarati et al. propose a timespace-network approach for the operational version of the LAP with a heterogeneous fleet and formulated the problem as a mixed integer linear program corresponding to a multi-commodity network flow problem with supplementary variables and constraints. The objective is the minimization of the total operational costs. They consider a week as time horizon, but the solution of very large instances impose to divide the time horizon into a set of rolling overlapping time windows of two or three days that involve fewer trains services (500 / 1000 each). Each time slice is then optimized using a branch-and-bound procedure in which the linear relaxations are solved with a DantzigWolfe decomposition. The solution of the problem for a slice determines the initial conditions for the following problem associated to the next slides. Computational experiments were conducted on real-life data (26 stations, 164 yards, 18 shops, 1988 train services, 1249 locomotives, 26 locomotive types). As in Florian et al. [1976], optimality was not reached, with gaps ranging from 3% to 7 %
To reduce the optimality gaps, Ziarati et al. [1999] strengthened the previous formulation with specific cutting planes, additional cuts that are based on the enumeration of feasible assignments of locomotive combinations to trains. They report an average reduction in integrality gap of about a third for problems of one, two, and three days time slice. The use of this cuts and of the new branching strategy (called branch-first, cut-second approach) consistently improve solution quality with modest increases in computing time.

An alternative approach to solve complex combinatorial problems has been proposed in Powell et al. [2001]; it is based on the approximate dynamic programming (ADP) framework.
The idea proposed by Powell et al. is to formulate the original problem as a

dynamic programming problem and solve, through ADP, a sequence of small sub-problems that can be managed optimally using commercial solvers (like CPLEX). This approach permits to deal with uncertainty in a general way allowing the modeling of a wide class of uncertainties even in complex real-life combinatorial problems.

The ADP framework has been extensively described in many papers (Marar and Powell [2009]; Marar et al. [2006]; Powell [2003]; Powell and Topaloglu [2003]; Powell et al. [2001, 2002, 2007]), technical reports (Powell and Bouzaiene-Ayari [2006]), conference proceedings (Powell and Bouzaiene-Ayari [2007]) and in a book (Powell [2007]).
The LAP is often formulated as a MIP problem, a class of problems which is treated for instance in Powell and Topaloglu [2005]; Powell et al. [2002]; Topaloglu and Powell [2006].

Moreover Powell et al. apply their approach to the solution of a real-life LAP. Focusing on recent application, in 2006 they start developing an application, sponsored by the Norfolk Southern Railroad and Burlington Northern Sante Fe Railroad. This application was claimed to solve the problem of assigning locomotives to trains over a planning horizon (a week for real-time planning, a month for strategic planning) capturing a high level of detail about both locomotives and trains, as well as a variety of complex business rules. Notably, the application simultaneously handled the problem of routing locomotives to shop location (maintenance centers). In 2007 this application was still in development in collaboration with Norfolk Southern Railroad.
Finally in 2009 the work of Powell et al. produced an application named Princeton locomotive and shop management system (PLASMA) which completed the user acceptance test at Norfolk Southern as a strategic planning system.

An important improvement in the realism of the LAP models has been provided in Ahuja et al. [2005b]. Ahuja et al. study a real-life locomotive scheduling faced by CSX Transportation Inc., a Class I U.S. railroad company. Following the requests of the CSX managers, who sponsored the research, Ahuja et al. focus on a weekly schedule and on the strategic version of the corresponding locomotive planning problem.
Ahuja et al. propose a Mixed Integer Programming (MIP) formulation, each

locomotive type correspond to a different commodity and the problem is modeled as a multicommodity flow with side constraints (the number of locomotives of each type is limited) on a space-time network where arcs denote trains and nodes denote events i.e. arrivals and departures of trains and locomotives (for a review of the network models and their application in locomotive and train scheduling see for instance Ahuja et al. [2005a]).
Defining the total cost as the sum of ownership, active, deadheading, light-traveling and consist-busting costs plus the penalty for the use of single-locomotive consists, the objective is to minimize the total costs finding active and deadheaded locomotives for each train, light-traveling locomotives and train-to-train connections.
Starting from the data provided by CSX, they consider an instance of the LAP with 538 trains running with different weekly frequencies, 119 stations and 5 types of locomotives. In the week the total number of trains (which differ at least for the running day) is 3324 and the resulting weekly space-time network consisted of 8,798 nodes (events) and 30,134 arcs (trips).
The proposed formulation does not consider some real-life constraints like the weekly consistency constraint (the same train running on different days should have the same locomotive assignment) and the train to train connection constraint (perform the same train to train connection for each pair of connected trains). Even without this constraint (which would increase dramatically the problem size), the MIP formulation consisted of 197,424 variables and 67,414 constraints and could not be solved to optimality or near-optimality using commercial software like CPLEX, even considering the linear programming relaxation of the problem. In order to deal with this large size instance, Ahuja et al. propose a decomposition-based heuristic approach that allows near-optimal solutions (using CPLEX) for real-life instances in moderate computing times and implicitly account for the consistency constraints. The first step of this heuristic approach transform the weekly scheduling problem in a daily scheduling one. This is done passing from the actual set of the weekly frequencies to the following binary set: cancel trains running less than 5 days a week (weekly frequency equal to zero) and set to 7 the frequency of the remaining trains (this simplification works because in the specific dataset provided by CSX the 94% of trains run 5 days a week or more). Even if the daily space-time network is significantly smaller, it contains 1,323 nodes and 30,034 arcs and finding an integer optimal solution is still very problematic. Ahuja et al. identify in the fixed-charge variables (fixed cost of deadheading and light-travelling) the principal

obstacle that prevent the daily problem to be solved to optimality or near-optimality. Then, the following three-step sequential heuristic approach is used to eliminate fixed-charge variables:

(i) Select among the admissible train to train connections the ones with lower impact on the cost function; the impact is assessed solving the linear programming relaxation of each problem obtained fixing connections one by one.

(ii) Identify a small but potentially useful set of light-travel arcs and, as for deadheading arcs, fix light-travel arcs one by one and select arcs relying on the impact on the cost functions

(iii) Once the fixed-charge variables are eliminated through the two previous steps, solve the integer program for the daily locomotive assignment without the fixed-charge variables obtaining a high-quality solution (in short time).

Ahuja et al. obtain a integer high quality solution for the daily scheduling problem in 15 minutes with CPLEX 7.0. The procedure is completed using this solution as the starting solution for a very large-scale neighborhood (VLSN) search algorithm that starting from this initial feasible solution repeatedly replaces it by an improved neighbor until we obtain a local optimal solution.

The solution of the daily problem is then heuristically adapted displacing locomotive from the fictitious trains to the actual trains (respectively inserted in the daily schedule and canceled from the weekly schedule) by the frequency quantization.

So a modified MIP flow formulation of the weekly problem is obtained from the solution of the daily problem resorting the original weekly frequency distribution. Anyway, this modified weekly problem still requires excessive computing time and so the corresponding multicommodity flow problem is heuristically converted into a sequence of single commodity flow problems with side constraints, one for each locomotive type. Finally, a VLSN search algorithm is applied to improve the feasible integer solution of the weekly locomotive scheduling problem obtained in the previous step.

Computational test were conducted on a real-life scenario: 3324 trains originating from and terminating at 119 stations and 3316 locomotives belonging to five locomotive types. The Algorithms made extensive use of CPLEX 7.0. and were tested on a Pentium III 750 MHz. The solution obtained in Ahuja

et al. [2005b] is substantially superior to the one provided by the software developed at CSX: the total cost is substantially reduced and the number of locomotives used dramatically decreases (by $350 \div 400$ units, depending on the scenario).
A technical document (Ahuja et al. [2006]) was also prepared to introduce possible extension of the model, e.g. CAB signal requirements, optimal routing of locomotive to fueling stations and shops to satisfy fueling and maintenance constraints. A more detailed presentation of these and other extensions can be found in Vaidyanathan et al. [2008a] where a generalized LAP is considered. Vaidyanathan et al. extended their previous model on several ways by incorporating in the planning problem all the real-world constraints needed to generate a fully implementable solution and by developing additional formulations necessary to transfer solutions of the models to practice. Vaidyanathan et al. propose also two alternative formulations for the generalized LAP: consist formulation, and hybrid formulation. The consist formulation is an extension of the locomotive flow formulation described in Ahuja et al. [2005b], which defines each locomotive type as a commodity and routes locomotives on the train network. In the consist flow formulation locomotive types are replaced by the consist types and each consist type is defined to be a single commodity and is routed on the train network. In locomotive flow formulation, single locomotives are assigned to trains and consist are the result of this assignment. In the consist flow formulation the solution is obtained starting from a set of consist already assembled. The optimal set of assembled consist is determined heuristically. The hybrid formulation allows the assignment of both assembled consist and single locomotives.
Vaidyanathan et al. point out that performances critically depends on the number and types of consists, as expected the greater the number of consists with different horsepower and tonnages, the better the quality of the solution.
The use of assembled consist restrict the solution space, this could lead, and leads, to a loss in optimality. Nevertheless, computational tests performed by Vaidyanathan et al. show that the optimal objective function value in the consist formulation may be just 5% higher than the one obtained in the locomotive flow formulation. The correct identification of the set of assembled consit is crucial to reduce as much as possible the optimality gap.
The (potentially) small optimality gap is highly compensated by many benefits:

(a) For some instances the locomotive flow formulation could not converge to a feasible solution in more than 10 hours of computation, while the consist flow formulation was able to optimally solve the same instances within a few minutes.
(b) The consist flow formulation use a much lesser number of constraints because allow to implicitly handle many constraints that have to be explicitly specified in the locomotive flow formulation; this well explains the shorter computation times and the rapid convergence toward an optimal solution.
(c) Railroads often impose complex rules on what locomotive types may be combined together, these requirements are very hard or impossible to impose in the locomotive flow formulation while are easy to enforce in the consist flow formulation.
(d) Consist-busting (and the corresponding cost) is reduced to a large extent

In fact, great improvements in solution speed and robustness, significant consist-busting reduction and easy implementation of complex constraints, make the consist flow formulation superior.
Some important real-life constraints cannot be inserted in the scheduling phase, so the models proposed in Ahuja et al. [2005b] and Vaidyanathan et al. [2008a] did not account for the fueling and maintenance feasibility of individual locomotive units. The fueling and maintenance constraints have to be imposed to specific locomotive units, not to locomotive types. This can be done in the locomotive routing phase, that follow the scheduling phase. Vaidyanathan et al. [2008b] developed methods that allow to route locomotive units on fueling and maintenance friendly routes while honoring the constraints seen in the scheduling phase.

4. Mathematical modeling

It is difficult to identify a representative mathematical model for the LAP. The LAP can be encountered in several contexts (rail freight, rail passenger, switch engines, in-plant railroad) and many different formulation are possible, depending on the problem type (single / multiple locomotive), on the objective function and on the constraints.
It seems a good idea to consider the more general problem (LAP with heterogeneous consists) since the other problems can be seen as simplified version of

the general one. In the OR literature, the most used formulation for the LAP with heterogeneous consists is the one used by Ahuja et al., a Mixed Integer Programming (MIP) formulation where each locomotive type correspond to a different commodity and the problem is modeled as a multicommodity flow with side constraints on a space-time network.

The model proposed by Ahuja et al. is also the more advanced under many aspects and may be considered the state of the art, hence it represents a good reference model. This model has been formulated in a new way in Vaidyanathan et al. [2008a] where authors introduce the consist flow formulation, in this new formulation each consist is defined as a commodity and the set of feasible consist represent the set of flowing commodities.
Vaidyanathan et al. define a space-time network $G = (N, A)$ where nodes N and arcs A are divided into different categories. Nodes belong to three different sets, the arrival nodes ($ArrNodes$) which represent the train arrival events, the departure nodes ($DepNodes$) which represent the departure events and the ground nodes ($GrNodes$) that allow the flow of consist from inbound trains to outgoing trains. The ($GrNodes$) allow to model easily train to train connection, light-travel and idling of consist in stations.

Arcs belong to four different sets, train arcs $TrArcs$ connect ($DepNodes$) and ($ArrNodes$), ground arcs $GrArcs$ connect ($GrNodes$) to ($GrNodes$) (train is idling in a station). Each arrival node $\in ArrNodes$ has a corresponding arrival ground node $\in GrNodes$, the same holds for departure nodes, connection arcs $CoArcs$ connect arrival nodes $\in ArrNodes$ to the corresponding arrival ground nodes $\in GrNodes$ and the same holds for departure nodes (these are the train to train connections). Finally light-traveling arcs $LiArcs$ connect ($GrNodes$) to ($GrNodes$) (train is light-traveling). The model assumes that the light-travel possibilities are given.

It is important to note that for each station the last ground node of the week is connected to the first ground node of the week of that station, through a ground arc such that the ending inventory of locomotives becomes the starting inventory in the next time period. This permits to count the locomotives used during the week, evaluating the flow of locomotives on arcs that cross the time line at midnight on Sunday (Sunday midnight is the check time, at this time there are no arrival or departure).

Each train l is characterized by the following attributes:
dep-time(l): the departure time of train l
arr-time(l): the arrival time of train l
dep-station(l): the departure station of train l
arr-station(l): the arrival station of train l
T_l: tonnage requirement for train l
β_l: HP per tonnage requirement for train l
E_l: the penalty for using single locomotive consist on train l

Given the set of all locomotive types K, k denotes a particular locomotive type belonging to K. Every k $\in$ K is characterized by the following attributes:
h^k: horsepower (HP) of a locomotive of type k
b^k: number of axles on a locomotive of type k
G^k: ownership cost of a locomotive of type k
B^k: fleet-size of a locomotive of type k
c_l^k: cost of assigning an active locomotive of type k to train l
d_l^k: cost of deadheading a locomotive of type k on train l
t_l^k: tonnage provided by a locomotive of type k to train l

Each train l has three sets of locomotives that could be assigned to it. MostPreferred[l] (locomotive types preferred), LessPreferred[l] (locomotive types accepted with a certain penalty) and Prohibited[l] (locomotive types not allowed).
C: set of consist types available for assignments whereas $c \in C$ denotes a specific consist type.
F_l: fixed cost for using a light arc l.
c_l^c:cost of assigning an active consist of type $c \in C$ to train arc l.
d_l^c: cost of assigning a deadheading consist, light-traveling consist or idling consist of type $c \in C$ if the train arc l belongs to the sets
$TrArcs$, $LiArcs$ or $CoArcs \bigcup GrArcs$ respectively.
α^{ck}: number of locomotives of type $k \in K$ in consist $c \in C$.
$I[i]$: set of arcs entering in the node i.
$O[i]$: set of arcs leaving the node i.
S: set of overnight arc crossing the Sunday midnight timeline.

The decision variables are the following

x_l^c: binary variable representing the number of active consist of type $c \in C$ on arc $l \in TrArcs$.
y_l^c: integer variable, representing the number of non-active consists (deadheading, ligh-traveling or idling) of type $c \in C$ on arc $\in AllArcs$ ($AllArcs = TrArcs \bigcup GrArcs \bigcup LiArcs \bigcup CoArcs$).
z_l: binary variable which takes value 1 if at least one consist flows on arc l and 0 otherwise.
z_c: binary variable which takes value 1 if consist type $c \in C$ is used and 0 otherwise.
s^k: integer variable that indicate the number of unused locomotives of type $k \in K$.

The constraints in the model are the following

- The constraint (2) imposes that to each train l is assigned exactly one active consist

- The constraint (3) imposes the locomotives flow upper bound on each train arc

- The constraint (4) ensures that the consists flow is balanced in every node for every consist type

- The constraint (5) imposes the locomotives flow upper bound on light arc

- The constraint (6) imposes that, for each locomotive class k, the number of used locomotives is no more than the available locomotives

- The constraints (7) and (8) ensure that the model extract a subset of p consist from the set C.

The weekly consist flow formulation for the LAP with a fixed number p of available consist types is the foolowing

$$min:\ w = \sum_{l \in TrArcs} \sum_{c \in C} c_l^c x_l^c + \sum_{l \in AllArcs} \sum_{c \in C} d_l^c y_l^c + \sum_{l \in LiArcs} F_l z_l - \sum_{k \in K} G^k s^k \quad (1)$$

$$subject\ to$$

$$\sum_{c \in C} x_l^c = 1 \quad (2)$$

$$\sum_{c \in C} \sum_{k \in K} \alpha^{ck}(x_l^c + y_l^c) \leq 12, \quad for\ all\ l \in TrArcs \quad (3)$$

$$\sum_{l \in I[i]} (x_l^c + y_l^c) = \sum_{l \in O[i]} (x_l^c + y_l^c), \quad for\ all\ i \in AllNodes,\ c \in C \quad (4)$$

$$\sum_{c \in C} \sum_{k \in K} \alpha^{ck}(y_l^c) \leq 12 z_l, \quad for\ all\ l \in LiArcs \quad (5)$$

$$\sum_{l \in S} \sum_{c \in C} \alpha^{ck}(x_l^c + y_l^c) + s^k = B^k, \quad for\ all\ k \in K \quad (6)$$

$$\sum_{l \in S} (x_l^c + y_l^c) \leq M z_c, \quad for\ all\ c \in C,\ M\ is\ a\ sufficiently\ large\ number \quad (7)$$

$$\sum_{c \in C} z_c = p \quad (8)$$

$$x_l^c \in 0,1, \quad for\ all\ l \in TrArcs,\ c \in C, \quad (9)$$

$$y_l^c \geq 0, \quad and\ integer \quad (10)$$

$$z_l \in 0,1, \quad for\ all\ l \in LiArcs \quad (11)$$

$$z_c \in 0,1, \quad for\ all\ c \in C \quad (12)$$

$$s^k \geq 0, \quad for\ all\ k \in K \quad (13)$$

5. The identification of the consist set

The locomotive fueling and maintenance constraints are typically considered in the routing phase, that follows the planning one. For this reason (and

also to not further complicate a very complex model) in Ahuja et al. [2005b] and Vaidyanathan et al. [2008a] the fueling and maintenance constraints are not inserted and are considered in a distinct paper that focuses on the freight locomotive routing problem (Vaidyanathan et al. [2008b]).
The fueling and maintenance constraint are relegated in the routing phase because the locomotive manager has to route the single specific locomotives (identified by their unique ID number) to fueling stations and shops to honor this constraints (it does not make sense to consider an entire class of locomotives in this case). Nevertheless, this does not necessarily imply that the planning phase has no role in the fueling and maintenance problem.
Another aspect usually not considered in the planning LAP is the robustness of the solution. In real applications uncertainties constantly affect the train schedules, specially in the management of freight train where delays (and locomotive disruptions) are more frequent and a part of trains work under a tonnage-base regime. It could be useful to identify a simple way that increase the robustness of the solution without introducing a robust optimization formulation of the problem (that would complicate an already complex model) and without introducing explicitly the uncertainties in the train schedules.
To deal with this two aspects, (fueling/maintenance constraints and robustness), is it possible to introduce in the model the concept of homogeneity, not considered in the previous models.
Fueling and maintenance constraints are considered first. The idea is to start from the planning phase inserting elements in the model that make easier the routing to fueling and maintenance stations in the following phase.
The profile of the set of potential consist C remains general, unspecified (every feasible consist type is a potential optimal consist) or is pre-defined by locomotive managers. The limitation of the number of feasible consists reduces optimization possibilities and could reduce the solution optimality. Nevertheless a judicious choice of available consist types could preserve optimality and simplify the model and the selection process since allows to honor complicates constraints that are difficult to insert in the model or make the model less tractable.
The idea is to take into account additional aspect, not considered so far in the optimal consist selection, such that the final planning LAP solution is easy to handle in the routing phase, where fueling and maintenance constraints are honored.
In CSX a locomotive should be sent to the shop every 92 days (John and Ahuja [2008]). So far, the consist are assembled without considering when a

locomotive becomes critical (i.e. the maintenance is scheduled within 7 days). Then the residual time to the next maintenance event (shortly *rtm* in the following) is in general different for each locomotive inside each consist. In other terms, consist are in general heterogeneous with respect to the *rtm* parameter. This fact has two important consequences in the routing phase:

(i) each heterogeneous consist must be busted in order to send the critical locomotives to the shop
(ii) critical locomotives are highly dispersed over many different stations over the entire network

As a consequence, we have a high number of consist busting firstly and a high number of locomotive travel toward the shops secondly, in other words high consist-busting costs, high travel costs, organizational and logistic complexity, increased risks for crews and equipment.
Building consist considering the *rtm* parameter permits to obtain homogeneous consist with the following positive impact in the routing phase:

(i) critical locomotives are grouped in critical consist that can be sent directly to shops avoiding the consist busting
(ii) in particular cases, the limited capacity of some shops could impose some consist busting, anyway their number is expected to be far lower than the one necessary with heterogeneous consist
(iii) critical locomotives are grouped in critical consist minimizing the number of stations where critical locomotives are located

With maintenance-homogeneous consists it is expected a significant reduction in consist busting and in travels of locomotive toward shops. Cost and risk could be reduced and maintenance logistic should be highly facilitated.

The routing phase may benefit from maintenance-homogeneity of consist also because it is (potentially) possible to assign critical consist to trains (or train sequences) that end their trip in stations equipped with a shop (or close to these stations). Is it also possible to exploit the maintenance-homogeneity in an alternative way. Locomotive failure is more frequent than expected. Critical consist may be considered more risky, they could be considered more prone to failure since more time is passed from the last maintenance. Then, the locomotive manager could decide to assign critical consist to trains less impacted by a disruption or which travel on tracks were recovery is more rapid and less costly.

Different parameters can be defined to evaluate the homogeneity of a consist. After having considered the *rtm* it is possible to consider the residual time to the next fueling event (*rtf*). If a consist is built using locomotives with very different ranges, the frequency of fueling events increases. On the contrary, a group of locomotives characterized by similar ranges needs to stop less frequently for fueling events. The parameters *rtm* and *rtf*, expressed respectively in days and hours, allow to identify maintenance-homogeneous and fueling-homogeneous consist types. The identification of homogeneous consist types is done preliminarily, before starting the optimization phase. In Vaidyanathan et al. [2008a] the set of potential consist types C was pre-specified, following the indications of CSX. What the optimization program do is just to identify the numerosity of the different consist types available in C. Obviously the expertise of locomotive managers cannot be substitute but could be integrated in this preliminary identification phase. So the identification phase helps to create the pre-specified consist set jointly with other company-specific constraints and allow to select consist types that are not captured by a simple cost optimization and that are more productive in the following phases (like routing). The identification phase is independent from the particular optimization strategy and can be associated to the optimization model proposed by Vaidyanathan et al. or to alternative models. The identification phase addresses the optimization specifying the set C. Reducing the set of potential consist types C could lead to a loss in optimality: costs can increase since more expensive consist type could be selected excluding some economic consist types.

To asses the impact of the identification phase on the optimality we should solve the optimization program under different scenarios considering different consist set C provided by different identification choices (as done in Vaidyanathan et al. [2008a]). Nevertheless we could make some initial considerations on the cost-benefit ratio of the identification phase analyzing the results obtained in Ahuja et al. [2005b] and in Vaidyanathan et al. [2008a]. According to CSX Corporation [2005] and CSX Corporation [2006] the CSX road freight locomotive fleet profile at December 2004 and at December 2005 were the following

Table 3: CSX road freight locomotive fleet profile in 2005, 2006 and 2011[a]

Locomotive class	Units 2005	Units 2006	Units 2011
AC4400CW	593	593	621
C40-8/C40-8W	532	532	529
SD40/SD40-2/SD40-3	404	402	529
SD70AC	220	220	20
SD50/SD50-2	177	177	174
AC6000CW	116	117	117
ES44DC	0	100	302
SD60I/SD60/SD60M	90	90	94
C44-9W	53	52	52
B40-8	32	32	50
SD70M	25	25	25
SD70AE	0	20	0
SD80AC	13	13	13
GP60	3	3	3
GP40/GP40-2	0	0	416
C39-8	9	0	0
C44-6W	1	0	0
ES44AC	0	0	300
GP38-2/GP38-2s	0	0	323
SD70MAC	0	0	220
Road Slug	0	0	190
Fleet size	2268	2376	3978

[a]data source for 2011: www.thedieselshop.us/CSX.HTML - (accessed April 21 2011)

The fleet data for 2011 were very detailed so to obtain a more compact table the classes with very low numerosity have been neglected in fleet 2011 (the reported fleet 2011 represent the 95.12% of the actual fleet 2011). The very similar locomotive classes have been aggregated (in fleet 2005 and 2006 the only aggregated classes are $SD60I/SD60/SD60M$).

According to CSX Corporation [2007], at December 2006 CSX operates 3,853 total locomotives, owned or leased long term, including 2,489 road and freight locomotives (65%), 1,175 yard and local service locomotives (30%) and 189 auxiliary units (used to produce extra traction for heavy trains in hilly terrain). The fleet received 200 high-horsepower units in 2007, in addition to the 100 locomotives added in 2006. These newer locomotives meet the Clean Air Acts emission standards and offer more fuel efficiency and greater reliability than earlier models. As a result, 150 leased units will be retired.
According to CSX Corporation [2009], at December 2009 CSX operates 4,071 total locomotives (of which 95% owned and 5% leased long term) including 3,539 road and freight locomotives (87%), 311 yard and local service locomotives (8%) and 221 auxiliary units. As of December 2009, 566 locomotives (14%) were held in temporary storage due to significant declines in volume. As volume returns, these locomotives could be placed back into service within a week, after restorative maintenance procedures are performed.

To asses the feasibility and the impact of the maintenance-homogeneity selection strategy on locomotive management it is useful to compare the locomotive availability in the CSX locomotive fleet with the locomotive utilization provided by Ahuja et al. [2005b].
The first paper (published in 2005) proposes the locomotive flow formulation and considers five locomotives types, the second one (published in 2008) considers six locomotive types and shows the superiority of the consist flow formulation. The first paper compare the results obtained by the optimization procedure developed by CSX and the one developed by Ahuja et al., table 4 resumes the results in term of used locomotives.

According to Ahuja et al. [2002] (from which Ahuja et al. [2005b] stems) the first scenario is characterized by 119 stations and 538 trains, each of which operate several days in a week such that there are 3324 weekly trains that differ for at least for the operating day (the corresponding weekly space-time network has 8798 nodes and 30134 arcs). Ahuja et al. [2005b] introduces two

Table 4: # of locomotives of different types - CSX vs Ahuja solution						
	Scenario 1		Scenario 2		Scenario 3	
Locomotive model	CSX	Ahuja	CSX	Ahuja	CSX	Ahuja
SD40	498	249	519	283	550	323
SD50	171	160	162	138	174	161
C40-8W	621	487	619	466	620	432
AC4400CW	164	154	155	154	155	155
AC6000CW	160	160	160	160	160	160
Total	1614	1210	1615	1201	1659	1231

additional scenarios (scenario 2 and 3) characterized by the same number of stations and trains but where trains differ in tonnage and horsepower requirements.
Vaidyanathan et al. [2008a] implement the consist flow formulation considering two smaller scenarios, scenario A (388 trains, 6 locomotive types, 87 stations) and scenario B (382 trains, 6 locomotive types, 87 stations) and provide the total number of locomotives used with 8 different sets C (from 3 consist types to 17 consist types). Table 3 resumes only the result for the consist flow formulation since the number of used locomotives is greater in the consist flow formulation (the corresponding solution costs is at least 5% greater than the one obtained by the locomotive flow formulation)

Table 5: total number of locomotives in the consist flow solution			
		# locomotives	
Consist set	# Consists	Scenario A	Scenario B
2[SD40], 3[SD40], 3[C40-8W]	3	1376	1388
& 2[AC6000CW], 2[C40-8W]∪1[SD40]	5	1330	1343
& 1[C40-8W]∪2[SD40],2[AC4400CW]	7	1183	1201
& 2[C40-8W], 2[SD40]∪1[SD60I]	9	1051	1064
& 2[C40-8W]∪[AC6000CW], 1[C40-8W]∪1[SD40]	11	1051	1064
& 1[SD60I]∪[GP40], 1[AC4400CW]∪1[SD40]	13	1045	1056
& 4[SD40], 2[GP40]	15	1047	1062
& 4[SD60I], 3[GP40]	15	1047	1063

The paper does not specify the number of locomotives detailed for each locomotive class but provides just the total number of locomotives.

According to Ahuja et al. [2002] the solution is obtained assuming an availability of 3316 locomotives belonging to five locomotive types. Nevertheless the number of 3316 locomotives (belonging to the five classes and available for freight trains) seems to be disproportionated (too big) with respect to the utilization of the locomotives provided in Ahuja et al. [2005b]. Even considering the less optimal case (the CSX solution requires 1659 locomotives in scenario 3) the percentage of used locomotives based on 3316 available locomotives, is very low and it is not realistic (49.97% of unuse locomotives). The problem is that if we consider the actual locomotive availability (fleets 2005, 2006 and 2001) the utilization of some locomotive classes exceeds the their actual availability because the problem was solved assuming 3316 available locomotives.

Table 6: Δ actual locomotive fleet vs scenario

	Locomotive fleet 2005		Locomotive fleet 2008		Locomotive fleet 2011	
Locomotive class	Ahuja	CSX	Ahuja	CSX	Ahuja	CSX
AC4400CW	428	419	438	429	466	457
AC6000CW	-44	-44	-43	-43	-43	-43
C40-8/C40-8W	45	-89	44	-90	42	-92
SD40-2	121	-146	119	-148	235	-32
SD50	16	6	16	6	-18	-28

It is clear that the actual availability of locomotives in 2005 and 2006 for the five models is far away from the one assumed in the paper. The considered scenarios are then realistic but does not depict the actual situation and the requests of locomotives are sized for a significantly larger fleet of 3316 units. This is evident considering the negative delta between the actual number of available locomotives in the class $AW6000AC$ and the locomotive requested in the scenarios. Nevertheless, selecting the first or the second scenario and the solutions provided by Ahuja we obtain a locomotive utilization which is feasible with both 2005 and 2006 fleets (with the exception of the class $AC6000AC$).
This solution is the result of a locomotive flow approach but (since Vaidyanathan et al. [2008a] provide only the total utilization for the consist flow solution)

to make an example we could assume that the locomotive utilization in scenario 1 is associated to a consist flow solution: in fact the consist flow solution differs only marginally from the locomotive flow solution (5%), then this assumption is not so wrong.
After a transition period we may assume that it is possible to group locomotives in subgroups each characterized by a specific maintenance date belonging to a specific week. The numerosity of each subgroup depends on the weekly capacity of shops. For the sake of simplicity we may guess that, since every locomotive has to be checked every 13 weeks, every year the number of maintenance events is four times the number of used locomotives. We suppose that the number of used locomotives provided by CSX in the three scenario is realistic and gives a a reference value of the shop maintenance capacity. Considering the higher and the lower values (1659 and 1614 locomotives) and a working year of 50 weeks we have that the maintenance weekly capacity of all the shops ranges in the interval $[133, 129]$ locomotives serviced every week. Assuming a weekly capacity of 134 locomotives and a weekly utilization of 1659 locomotives, it is possible to group locomotives in 13 groups, each of which is is rtm-homogeneous and is composed by 128 locomotives. Clearly the rtm for the 13 groups range from 91 to 1, and the group with $rtm \in [1, 7]$ is the one composed by critical consist.
In this scenario, every week we need to replace 128 locomotives (the critical locomotives) that are routed toward the shop in the incoming week. Since locomotive managers are interested in cyclic locomotive scheduling, the set of consist formed by the 128 critical locomotives, contains the same consist types (i.e. presents the same distribution of critical locomotives) every week. In other words, the distribution of locomotive classes inside the critical group should be the same observed in the set of used locomotives. This means that the locomotive class with the lowest (positive) delta between availability and utilization is determinant for the feasibility of the maintenance-homogeneity selection strategy.
Then the class $SD50$ has the smaller positive delta, the residual available locomotives amount to 16 units. This value has to be greater than the one used to replace the critical locomotives belonging to the class $SD50$. As said the number of critical locomotives belonging to the class $SD50$ is given by $\frac{177}{13} = 13.61$, where 177 is the numerosity of the class (see table 3).
Since $16 > 13.61$, the maintenance-homogeneity selection strategy is feasible for this scenario. A similar result is obtained considering the percentage weight of the class $SD50$ in the fleet 2004 (7.84%) and assuming that the 128

critical locomotives leaving the set of active locomotives are distributed as the used locomotives, then $16 > 0.84 \cdot 128 = 10.03$. This is just an example and the small margin (only 2 locomotives more than 13.61) does not protect from disruptions in the locomotive class $SD50$, anyway, as stressed, the considered utilization is suited for a much larger fleet size (3316 locomotives instead of 2258). So the maintenance-homogeneity selection strategy should be feasible (eventually adopting some minor changes in the locomotive availability) and does not interfere with the optimality search and the ratio cost-benefit seems to be interestig, specially in view of the following phases (like routing).

To asses the impact of the fueling-homogeneity selection strategy we need much more data. The range depends on the fuel capacity of the locomotive and on the fuel consumption rate. Assuming a planar and straight track (and an average value of the adhesion coefficient between rail and wheel) the fuel consumption depends on the HP provided by the locomotive which depends by the tonnage of the train and the velocity at which the train should be pulled.
The maintenance-homogeneity selection has a direct impact on the composition of the set C because we have to guarantee the availability of the correct locomotive types. However it does not impose preliminary restrictions on the setting up of the consist, every locomotive model could be joined with any other locomotive model to form a consist. In the worst case we may temporarily form consist that are not strictly homogeneous (for instance, one locomotive is critical, the others becomes critical within 3 days) and you can accept, paying some penalty, to send a locomotive to the shops some days before (or after) the scheduled maintenance day. On the contrary the fueling-homogeneity selection strategy heavily affect the consist set C since impose some constraints on the available consist types, for instance: the class $AC6000$ and the class $SD40$ could have very different fuel consumption rates and this should exclude the possibility to form a consist joining locomotive belonging to this two classes.
In fact the fueling-homogeneity selection strategy appear to be costly and to evaluate the benefits of its adoption we should consider some real scenarios that permit to evaluate how may fueling events is it possible to save with this strategy. As said, the consumption depends on the train tonnage and on the train speed, this are parameters that essentially depend on the train type (Auto, Intermodal, Merchandise). According to GE Harris Energy Systems [2000] approximately 5% of a total Class I road fleet would run out of fuel

during the course of a year even if, on average, Class I locomotive tanks are refueled when they are 60% full, so the number of fueling event is higher than the one we should expect looking at the locomotive fuel capacity. For this reason it seems valuable to evaluate the feasibility of the (more costly) fueling-homogeneity selection strategy.

It is important to note that the initial set C could be obtained as a solution of an optimization problem (different from the locomotive assignment and simpler) and that both maintenance-homogeneity and fueling-homogeneity selection strategy can be imposed introducing suitable constraints. A soft constraints that associate a penalty proportional to the heterogeneity of the consist in terms of rtm seems to be suitable for the maintenance-homogeneity selection. On the contrary a hard constraints that fix the maximum range difference in terms of rtf between two locomotive classes, seem to be the right choice for the fueling-homogeneity selection of the consists. In this preliminary optimization problem (selection phase) the objective function should consider only active costs because they are dominant and because the choice of the consist type is done looking at the demand of the active trains. The problem becomes more simple since the number of variables and constraints is reduced and the problematic fixed-charge variables disappear (alternatively these constraints could play a role directly in the the locomotive assignment model).
Neglecting the deadhead, ligh-travelling and ownership cost we loose sensibility on the cost optimality but this is not the crucial point in this initial phase because we are defining the set of potential consists that should provide benefits in terms of maintenance routing, fueling routing and robustness that are not evaluable in the planning phase looking strictly at the costs.

To promote the robustness of the solution we can still look at the homogeneity but this time it is not the consist internal homogeneity.
A robust solution should be less sensitive to disruptions and delays and should allow for an easier reaction to this kind of unexpected events. If it would be possible to use the same type of consist to pull all the trains, this solution would make a big step toward robustness. Indeed this ideal situation is associated to the highest possible consist fleet homogeneity: every consist disruption or delay could be absorbed in an easier way since any element of the consist set could replace any disrupted or delayed consist.
Clearly it is impossible to achieve this fleet homogeneity without incurring

in an extreme increase in costs, since in this case the powerful consist would be used for every train. Moreover there are some technical constraints that prevent the use of the same type of consist for every train.
Vaidyanathan et al. [2008a] introduce in their model two additional constraints that impose a limit on the number of consist types. The idea is to improve substitution opportunities finding a small set of consist types that offers a good solution as a large one. This approach offer good results, Vaidyanathan et al. provide 8 scenarios with a number of consist types ranging from 3 to 17 (3, 5,. . ., 15, 17) and show that these additional constraints ensure that the model identify a set of p optimal consist types. Vaidyanathan et al. consider a scenario with 388 trains, 6 locomotive types and 87 stations and show that increasing the number p from 3 to 9, the objective function value (the total cost) is improved by more than 17% whereas increasing p from 9 to 17 the costs is further reduced by just a 3%.
Instead of fixing the number of consist types introducing two additional constraints in the optimization model, we can alternatively impose the a limit in the number of available consists types in the selection phase (the preliminary optimization phase). To limit the number of consist type we can integrate the preliminary optimization problem (selection phase) as follow. We determine the set of potential consist C without considering any (fleet or consist) homogeneity constraints. Then as done by Vaidyanathan et al. [2008a] we define a binary variable z_c that assume value 1 if a consist type is keep inside the set C and value 0 if the consist is eliminated from C.
Then we define the sum

$$Z = \sum_{c \in C} z_c \tag{14}$$

and the active cost function

$$\sum_{l \in TrArcs} \sum_{c \in C} c_l^c x_l^c \tag{15}$$

and we consider the minimization of a convex combination of these two terms

$$min: \ w = \Big(\sum_{l \in TrArcs} \sum_{c \in C} c_l^c x_l^c \Big) \eta + \Big(\sum_{c \in C} z_c \Big) (1 - \eta), \quad \eta \in [0, 1] \tag{16}$$

The user can define the weight of the two terms obtaining more homogeneous but costly consist fleets or less robust but economic solutions.

$$min: \ w = \Big(\sum_{l \in TrArcs} \sum_{c \in C} c_l^c x_l^c \Big) \eta + \Big(\sum_{c \in C} z_c \Big) (1 - \eta) + \lambda \Big(\sum_{c \in C} \Delta_{rtm}^c \Big) \tag{17}$$

$$\eta \in [0,1], \ \lambda \geq 0 \tag{18}$$

subject to

$$\Delta_{rtf}^c \leq \Delta_{rtf}^{max} \quad for \ all \ c \in C \tag{19}$$

$$\sum_{c \in C} x_l^c = 1 \tag{20}$$

$$\sum_{l \in I[i]} x_l^c = \sum_{l \in O[i]} x_l^c, \quad for \ all \ i \in AllNodes, \ c \in C \tag{21}$$

$$\sum_{l \in S} \sum_{c \in C} \alpha^{ck}(x_l^c) + s^k = B^k, \quad for \ all \ k \in K \tag{22}$$

$$x_l^c \in 0,1, \quad for \ all \ l \in TrArcs, \ c \in C, \tag{23}$$

$$z_c \in 0,1, \quad for \ all \ c \in C \tag{24}$$

$$s^k \geq 0, \quad for \ all \ k \in K \tag{25}$$

Conclusions and future work

We describe the introduction of a preliminary optimization program (selection phase) that determines the set of the consist types initially available for the solution of the LAP. This phase could identify consist types that are not captured by a purely cost-oriented selection but that can be very useful specially in the routing phase, where they could simplify maintenance routing and fueling routing, producing saves that should not be achieved using (apparently) more economic consist types. The selection phase may be implemented as a simpler minimization program since the objective consider only active costs. We describe an alternative way to promote consist fleet homogeneity (and so solution robustness) expressing the objective function

in the preliminary minimization program as a convex combination of the active cost function and the total number of consist available in C.
Future researches should asses in a quantitative way the ratio costs-benefits for both the maintenance-homogeneity and the fueling-homogeneity strategies. It is also interesting the quantitative evaluation of the cost of robustness when consist types homogeneity is imposed in the preliminary phase instead of directly in the model as done in Vaidyanathan et al. [2008a].
The final objective is to integrate the preliminary phase with the solution of the LAP. The formulation proposed by Vaidyanathan et al. [2008a] could be the starting point for the solution of the LAP and alternative aggregations and heuristic procedures are possible. Alternatively, it seems promising to reformulate the problem as a set partitioning problem and to exploit the dominant number of variables over the number of constraint adopting a solution strategy based on the column generation approach.

References

R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network flows: theory, algorithms, and applications*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

R. K. Ahuja, J. Liu, J. B. Orlin, D. Sharma, and L. A. Shughart. Solving real-life locomotive-scheduling problems. Tech report, MIT Sloan School of Management, 2002. URL `http://dspace.mit.edu/bitstream/handle/1721.1/1799/4389-02.pdf?sequence=2`. Accessed Mar. 31, 2011.

R. K. Ahuja, C. B. Cunha, and G. Sahin. Network models in railroad planning and scheduling. *TutORials in operations research*, 1:54–101, 2005a.

R. K. Ahuja, J. Liu, J. B. Orlin, D. Sharma, and L. A. Shughart. Solving real-life locomotive-scheduling problems. *Transportation Science*, 39:503–517, 2005b.

R. K. Ahuja, L. A. Shughart, J. Liu, and B. Vaidyanathan. An optimization-based approach for locomotive planning. Tech report, Innovative Scheduling, 2006. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.7865&rep=rep1&type=pdf`. Accessed Mar. 31, 2011.

A. Baceler and A. Garcia. An optimization approach to the locomotive scheduling problem in Brazilian railways. In R. Ravitharan, editor, *Rail Achieving Growth: Conference Proceedings*, Melbourne, 2006. RTSA.

J. M. P. Booler. The solution of a railway locomotive scheduling problem. *Journal of the Operational Research Society*, 31:943–948, 1980.

J. Bramel and D. Simchi-Levi. Probabilistic analyses and practical algorithms for the vehicle routing problem with time windows. *Operations Research*, 44:501–509, 1996.

U. Brannlund, P. O. Lindberg, A. Nou, and J. E. Nilsson. Railway timetabling using Lagrangian relaxation. *Transportation Science*, 32:358–369, 1998.

Cambridge Systematics Inc. 2004 Freight rail component of the Florida rail plan. Tech report, Florida Department of Transportation, 2005. URL `www.dot.state.fl.us/rail/Publications/Plans/2004/FreightComponentFull.pdf`. Accessed Mar. 31, 2011.

A. Charnes and M. H. Miller. A model for the optimal programming of railway freight train movements. *Management Science*, 3:74–92, 1956.

J. F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32:380–404, 1998.

J. F. Cordeau, F. Soumis, and J. Desrosiers. A benders decomposition approach for the locomotive and car assignment problem. *Transportation Science*, 34:133–149, 2000.

J. F. Cordeau, G. Desaulniers, N. Lingaya, F. Soumis, and J. Desrosiers. Simultaneous locomotive and car assignment at VIA Rail Canada. *Transportation Research Part B: Methodological*, 35:767–787, 2001.

CSX Corporation. Csx corporation 2004 corporate profile. Tech report, CSX Corporation, 2005. URL `http://idc.api.edgar-online.com/EFX_dll/EDGARpro.dll?FetchFilingHtmlSection1?SectionID=3847768-5067-127634&SessionID=FwUbHHC7QJSIcP7`. Accessed Apr. 30, 2011.

CSX Corporation. Csx corporation 2005 corporate profile. Tech report, CSX Corporation, 2006. URL `http://idc.api.edgar-online.com/efx_dll/edgarpro.dll?FetchFilingConvPDF1?SessionID=wx5cHZteg0JJ5qf&ID=4483224`. Accessed Apr. 30, 2011.

CSX Corporation. Csx corporation 2006-2007 corporate profile. Tech report, CSX Corporation, 2007. URL http://edgar.brand.edgar-online.com/EFX_dll/EDGARpro.dll?FetchFilingHTML1?ID=5413635&SessionID=rVuSHq-zAfhVxA7. Accessed Apr. 30, 2011.

CSX Corporation. Csx corporation 2009 annual report. Tech report, CSX Corporation, 2009. URL http://www.annualreports.com/HostedData/AnnualReports/PDFArchive/csx2009.pdf. Accessed Apr. 30, 2011.

M. Florian, G. Bushell, J. Ferland, G. Guerin, and L. Nastansky. The engine scheduling problem in a railway network. *INFOR*, 14:121–138, 1976.

M. A. Forbes, J. N. Holt, and A. M. Watts. Exact solution of locomotive scheduling problems. *Journal of the Operational Research Society*, 42:825–831, 1991.

A. Fügenschuh, H. Homfeld, A. Huck, and A. Martin. Locomotive and wagon scheduling in freight transport. In R. Jacob and M. Müller-Hannemann, editors, *ATMOS 2006 - 6th Workshop on Algorithmic Methods and Models for Optimization of Railways*, Zürich, 2006. Internationales Begegnungs und Forschungszentrum für Informatik (IBFI).

A. Fugenschuh, H. Homfeld, A. Huck, A. Martin, and Z. Yuan. Scheduling locomotives and car transfers in freight transport. *Transportation Science*, 42:478, 2008.

GE Harris Energy Systems. RSSI locomotive tracking white paper. Tech report, GE Harris Energy Systems, 2000. URL http://www.arema.org/files/library/1999_Conference_Proceedings/00051.pdf. Accessed Apr. 30, 2011.

L. Grenzeback, D. Hunt, J. Bryan, B. Muten, D. Simpson, and D. Jacoby. Rail freight transportation bottom line report. Tech report, American Association of State Transportation and Highway Officials, 2008. URL http://www.dot.state.oh.us/. Accessed Mar. 14, 2011.

T. Illés, M. Makai, and Z. Vaik. Railway engine assignment models based on combinatorial and integer programming - operations research report 2005-04. Tech report, Eötvös Loránd University of Sciences - Operations Research Department, 2005. URL http://www.cs.elte.hu/opres/orr/download/ORR05_4.pdf. Accessed Mar. 31, 2011.

T. Illés, M. Makai, and Z. Vaik. Combinatorial optimization model for railway engine assignment problem. In L. G. Kroon and R. H. Möhring, editors, *5th Workshop on Algorithmic Methods and Models for Optimization of Railways*. Internationales Begegnungs und Forschungszentrum für Informatik (IBFI), 2006.

P. Ireland, R. Case, J. Fallis, C. Van Dyke, J. Kuehn, and M. Meketon. The Canadian Pacific Railway transforms operations by using models to develop its operating plans. *Interfaces*, 34:5–14, 2004.

A. John and R. K. Ahuja. Locomotive models at CSX Transportation. Tech report, Innovative Scheduling, 2008. URL `http://www.innovativescheduling.com/Files/Presentations/ConferencePresentation3.pdf`. Accessed Apr. 30, 2011.

N. Lingaya, J. F. Cordeau, G. Desaulniers, J. Desrosiers, and F. Soumis. Operational car assignment at VIA Rail Canada. *Transportation Research Part B: Methodological*, 36:755–778, 2002.

M. E. Lübbecke and U. T. Zimmermann. Engine routing and scheduling at industrial in-plant railroads. *Transportation Science*, 37:183–197, 2003.

A. Marar and W. B. Powell. Capturing incomplete information in resource allocation problems through numerical patterns. *European Journal of Operational Research*, 197:50–58, 2009.

A. Marar, W. B. Powell, and S. Kulkarni. Combining cost-based and rule-based knowledge in complex resource allocation problems. *IIE (Institute of Industrial Engineers) Transactions*, 38:159–172, 2006.

G. Maróti and L. Kroon. Maintenance routing for train units: the transition model. *Transportation Science*, 39:518–525, 2005.

D. H. Noble, M. Al-Amin, and R. G. J. Mills. Production of locomotive rosters for a multi-class multi-locomotive problem. *Journal of the Operational Research Society*, 52:1191–1200, 2001.

B. Paoletti and S. Cappelletti. *The optimization model for locomotive rostering in the fleet planning project*, pages 362–365. Institute of Electrical and Electronics Engineers - IEEE, Beijing, Cina, 2007. URL `http://`

ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4281679. Accessed Mar. 31, 2011.

W. B. Powell. Dynamic models of transportation operations. *Handbooks in Operations Research and Management Science*, 11:677–756, 2003.

W. B. Powell. *Approximate dynamic programming: solving the curses of dimensionality*. Wiley-Interscience, New York, 2007.

W. B. Powell and B. Bouzaiene-Ayari. Approximate dynamic programming for locomotive optimization. Tech report, Princeton University, 2006. URL http://academic.research.microsoft.com/Paper/4518847. Accessed Mar. 31, 2011.

W. B. Powell and B. Bouzaiene-Ayari. Approximate dynamic programming for rail operations. In C. Liebchen, R. K. Ahuja, and J. A. Mesa, editors, *ATMOS 2007 - 7th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems*. Internationales Begegnungs und Forschungszentrum für Informatik (IBFI), 2007.

W. B. Powell and H. Topaloglu. Stochastic programming in transportation and logistics. *Handbooks in Operations Research and Management Science*, 10:555–636, 2003.

W. B. Powell and H. Topaloglu. Fleet management. *Applications of stochastic programming*, page 185215, 2005.

W. B. Powell, J. A. Shapiro, and H. P. Simo. A representational paradigm for dynamic resource transformation problems. *Annals of Operations Research*, 104:231–279, 2001.

W. B. Powell, J. A. Shapiro, and H. P. Simo. An adaptive dynamic programming algorithm for the heterogeneous resource allocation problem. *Transportation Science*, 36:231–249, 2002.

W. B. Powell, B. Bouzaiene-Ayari, and H. P. Simo. Dynamic models for freight transportation. *Transportation*, 14:285–365, 2007.

K. V. Ramani. An information system for allocating coach stock on indian railways. *Interfaces*, 11:44–51, 1981.

J. A. Sabino, J. E. Leal, T. Stützle, and M. Birattari. A multi-objective ant colony optimization method applied to switch engine scheduling in railroad yards. *Pesquisa Operacional*, 30:486–514, 2010.

V. Scholz. Knowledge-based locomotive planning for the Swedish Railway. Tech report, Swedish Institute of Computer Science (SICS), 2000. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.7764`. Accessed Mar. 31, 2011.

L. S. Thompson. *Conclusions: competitive tendering in railways - what can we learn from experience?*, volume 2007, pages 209–245. OECD (Organisation for Economic Co-operation and Development) Publishing, Paris, 2007.

H. Topaloglu and W. B. Powell. Dynamic-programming approximations for stochastic time-staged integer multicommodity-flow problems. *INFORMS Journal on Computing*, 18:31, 2006.

B. Vaidyanathan, R. K. Ahuja, J. Liu, and L. A. Shughart. Real-life locomotive planning: new formulations and computational results. *Transportation Research Part B: Methodological*, 42:147–168, 2008a.

B. Vaidyanathan, R. K. Ahuja, and J. B. Orlin. The locomotive routing problem. *Transportation Science*, 42:492–507, 2008b.

M. B. Wright. Applying stochastic algorithms to a locomotive scheduling problem. *Journal of the Operational Research Society*, 40:187–192, 1989.

K. Ziarati, F. Soumis, J. Desrosiers, S. Gelinas, and A. Saintonge. Locomotive assignment with heterogeneous consists at CN North America. *European journal of operational research*, 97:281–292, 1997.

K. Ziarati, F. Soumis, J. Desrosiers, and M. M. Solomon. A branch-first, cut-second approach for locomotive assignment. *Management Science*, 45: 1156–1168, 1999.

K. Ziarati, H. Chizari, and A. Nezhad. Locomotive optimization using artificial intelligence approach. *Iranian Journal of Science & Technology, Transaction B, Engineering*, 29:93–105, 2005.