Optimization of Locomotive Management and Fuel Consumption in Rail Freight Transport

Huaining Tian

A Thesis
In the Department
of
Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Computer Science) at
Concordia University
Montreal, Quebec, Canada

December 2017

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Huaining Tian

Entitled: Optimization of Locomotive Management and Fuel Consumption in Rail Freight Transport

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Computer Science)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________ Chair
Dr. Anjali Awasthi

______________________ External Examiner
Dr. Jean-François Cordeau

______________________ External to Program
Dr. Kudret Demirli

______________________ Examiner
Dr. Thomas Fevens

______________________ Examiner
Dr. Jaroslav Opatrny

______________________ Thesis Supervisor (s)
Dr. Brigitte Jaumard

____

Approved by ______________________
Dr. Sudhir Mudur Chair of Department or Graduate Program Director

Jan. 24, 2018
Date of Defence

____
Dr. Amir Asif Dean, Faculty of Engineering and Computer Science

Abstract

Optimization of Locomotive Management and Fuel Consumption in Rail Freight Transport

Huaining Tian, Ph.D.
Concordia University, 2017

For the enormous capital investment and high operation expense of locomotives, the locomotive management/assignment and fuel consumption are two of the most important areas for railway industry, especially in freight train transportation.

Several research works have been done for the Locomotive Assignment Problem (LAP), including exact mathematics models, approximate dynamic programming and heuristics. These previously published optimization works suffer from either scalability or solution accuracy issues. In addition, each of the optimization models lacks part of the constraints that are necessary in real-world locomotive operation, e.g., maintenance/shop constraints or consist busting avoidance. Furthermore, there are rarely research works for the reduction of total train energy consumption on the locomotive assignment level.

The thesis is organized around three main contributions. Firstly we propose a consist travel plan based LAP optimization model, which includes all the required meaningful constraints and which can efficiently be solved using large scale optimization techniques, namely column generation (CG) decomposition. Our LAP model uses the number of consist travel plans to evaluate the occurrence of consist busting.

In addition, a new column generation acceleration architecture is developed, that allows the subproblem, i.e., column generator to create multiple columns in each iteration, with each column being an optimal solution for a reduced sub-network. This new CG architecture greatly reduces the computational time comparing to our original LAP model.

For train fuel consumption, we derive, linearize and integrate a train fuel consumption model into our LAP model. To avoid potential train collision by train rescheduling that the new model requires, I establish a conflict-free pre-process to assign each train reasonable time windows. The new LAP-fuel consumption model works fine for the optimization of the train energy exhaustion on the locomotive assignment level.

For the optimization models above, the numerical results are conducted on the railway network infrastructure of Canada Pacific Railway (CPR), with up to 1,750 trains and 9 types of locomotives over a two-week time period in the entire CPR railway network.

Acknowledgments

I wish to express my gratitude to my supervisors, Dr. Jaumard. Her encouragement, guidance, and deep support during my study enabled me to develop an understanding of the subject. Her wide knowledge, detailed and constructive comments have been a great value for me. Her timely financial support throughout my PhD helped me to concentrate on my work.

I am grateful to all my student colleagues for their help and support. I would like to specially thank Mr. Peter Finnie of Canadian Pacific Railway, that supports me so much in our cooperation works in my PhD research career.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Railway industry transports goods and/or passengers by trains over tracks. Nowadays railroad transportation becomes a safe, high energy-efficient, and environment-friendly means of conveyance, especially for bulk goods [25].

1.1.1 Railway Industry in North America

The railway industry in North America has been developed over almost 200 years [1]. Nowadays, the railway transportation still plays an important role in modern logistic system and the economy in both United States and Canada, especially for freight train service. Demand for freight trains will continue to increase significantly in the future: by year 2045 from 2015, the goods transported by freight rail services is expected to increase by 24% by weight or 82% by value [7].

In Canada, 75% of ground-transported goods are by railway in tonne-miles [2]. There are two Class I transcontinental railways: Canadian National Railway (CN) and Canadian Pacific Railway (CPR), which are also two of the most efficient and profitable railways in North America. From the annual report 2016 from Transport Canada [8], Class I railway carriers operate 2,700 locomotives running with 51,600 freight cars over 45,199 route-kilometers (km) of track. Most of the rails and equipments belongs to CN and CPR.

For the railway companies, the energy expense is a tremendous part of the operation. Based on CPR Annual Report 2016 [6], fuel expense in 2016 is $567 million, which is 15% of the total annual operation expense for CPR. Suppose the fuel price raises in the future as it did in the past years, the energy expense would be even more conspicuous, e.g., from the same document, in 2014, the fuel expense of CPR was almost doubled to $1,048 million

(24% of the total operation expense).

Depending on the jurisdiction, the railway companies face different challenges, e.g., in Europe the just-in-time operation is considered to be the most important criterion to manage passenger trains. On the other hand, for North American railway companies, the scheduling of freight trains to meet customer demand is considered a more important challenge. In general, the freight trains have to share the railway network with passenger trains. In North America, freight trains have priority over passenger trains.

1.1.2 Railway Challenges

Although rail transportation has many advantages and been developed for long time, it still has some disadvantages. The main drawback for railway industry is that it is capital-intensive, including not only the construction of rail tracks, but also the investment on the purchase (or lease) & maintenance of locomotives and cars. For example purchase of a new locomotive means a substantial amount of capital investment (around $5 Million) without instant profit. In addition, these train/locomotive fuel consumption listed in previous section is also a high expense. These disadvantages lead to the main purpose of this thesis: optimization of locomotive management and fuel consumption in freight train transportation.

This thesis focuses on freight trains over passenger trains because of their different characteristics. The configurations of freight trains, which includes multiple, different types of locomotives with full or empty cars, are more complex than passenger trains. A passenger train usually has only one or two locomotives of at most two types, with two types of first-class cars and also two types of second-class cars. For the limited types of these equipments, usually we can solve the assignment problem of both locomotives and cars to passenger trains within one problem, but for freight trains they are two separated assignment problems. Even only considering the configuration and assignment of locomotives for freight trains, the problem is more complex. A freight train can be led by from one to six locomotives, with nine different types, that means the potential number of configurations will be up to 9^6.

The operation of passenger trains usually is cyclic and more time-sensitive so it prefers a fixed time schedule, whereas freight trains can be operated with a flexible schedule within reasonable time windows. That leads to the potential to adjust the speed of trains, so to reduce the fuel consumption. In Section 1.2 we will provide details of train rescheduling with time windows.

In addition, in freight transportation the flow of each type of good usually is only one direction. For example, in Canada, the grain or potash is only transported from central area to the coasts, the imported Asian cars are only from West coast to the East area. Since in most cases, the freight cars for each type of goods are unique, they have to be transferred back empty to their original position for being reused. Backward trains of empty cars need much less locomotive power than the full onward ones. It is obvious that using the same configuration of locomotives is a waste of locomotive usage. However, the locomotives of extra unused power can be reassigned to other trains for more efficient utilization, so that the total number of required locomotives to satisfy the given trains' requirement will be reduced. The locomotive assignment will obtain additional benefit if we can reschedule some trains. For example, for an outbound train A, the locomotives it needs have to be relocated from other station to guarantee its schedule. Now we can just delay it until inbound train B comes and re-assign its locomotives to train A. In conclusion, comparing to the time-sensitive passenger train operation with limited train configurations, the locomotive management for freight trains has more flexibility (as well as additional complexity), and the optimization has the potential to gain enormous benefit.

1.2 Problem Statement

1.2.1 Railway Generalities

Before discussing the detail of the locomotive management and fuel consumption problems, we introduce some generalities that will be used throughout the thesis.

A train is composed of rail cars which are pulled by locomotives. A freight railway company usually has several types of locomotives with different attributes: horsepower, tractive effort, brake type, axle number, fuel consumption rate, operation cost, etc. Usually more than one locomotive is used to pull a train. This group of locomotives, which works together to pull a train, is called *consist*.

As a result of the routing and scheduling process, a locomotive can be *active* (i.e., pulling a train), *deadheading* or *deadhauling*. Deadheading locomotives operate under their own power but do not pull trains. Deadhauling locomotives are not pulling and their engines are not operating. We will not distinguish deadheading from deadhauling locomotives. Because the weight of one locomotive is only around 2% of the total weight of the train, so the effect whether its engine on or off on the fuel consumption is very small. The deadheading operation is profitless but is an important method to relocate the locomotives that makes the usage/arrangement of locomotives more flexible.

In locomotive management, there is a time-consuming process called *consist busting*.

Consist busting means disassembling the consist of an inbound train into stand alone locomotives and reassigning them to several outbound trains. Consist busting means additional labor & operational cost and time requirements. It also reduces the robustness of the train schedule because it allows an outbound train to get locomotives from multiple inbound trains. If any of the source inbound trains is delayed, the outbound train has to be delayed as well. So consist busting should be avoided as much as possible.

The locomotives should undergo regular maintenance examinations in a shop station, typically every three months. A locomotive due to maintenance in the planning period under study is called *critical* locomotive. In order to follow the US legislation, a locomotive which misses its regular maintenance has to be turned off and towed to the shop. To balance the early shopping cost and the risk of late/failed shopping process is another issue for the locomotive management.

The locomotive consumes fuel in operation for the train schedule. The fuel consumption rate for a locomotive is affected by the engine type and horsepower output. Since the speed of train depends on the power of its consist, its characteristics, e.g., weight, number of cars, *etc.,* and geographical factors, the total fuel consumption of a train on certain segments of rail is a complex function of all the influential factors.

1.2.2 Locomotive Assignment Problem

The locomotive assignment problem (LAP) deals with the fleet of locomotives and the assignment of those locomotives to scheduled trains. The optimization of LAP is to minimize the total number and/or cost of assigning enough locomotives to given trains while all the technical, industrial and business constraints are satisfied. Typical constraints are horsepower, locomotive maintenance, minimum number for safety, *etc.*.

Usually for LAP, the given train schedule is fixed and untouchable. Yet it will be beneficial if we can change some of the train schedule, e.g., delay train A for an acceptable time to wait train B which will provide its locomotive(s) instead of deadhead extra locomotives from remote. However, the train schedule is less flexible and more complex than road and air transportation for its hard constraints, e.g., scheduled trains meeting over given sidings. So the potential train rescheduling algorithm should not either interfere too much with other trains operation or change the train meet events.

1.2.3 Locomotive/Train Fuel Consumption

The flexibility of train rescheduling benefits not only locomotive assignment, but also train energy consumption by speed adjustment. That brings another optimization: train fuel consumption problem, i.e., a mathematical model to minimize fuel consumption by controlling the train schedule and/or speed, which will also interact with the locomotive assignment problem. The key point is to find a precise and linearizable train fuel consumption model and integrate it to LAP model.

This thesis targets locomotive/train management in freight train area, including locomotive assignment problem and the energy consumption for given train schedule. For LAP we first propose an optimization model (marked as LAP-Orig) with CG decomposition that works with all needed constraints, e.g., locomotive maintenance constraint, and with consideration of consist busting avoidance. Furthermore, we develop a new CG architecture that allows pricing problem to generate multiple potential optimal results for each iteration, and extend LAP-Orig into multi-column LAP model (marked as LAP-Multi-CG). Finally we create a collision-free pre-processing algorithm for train rescheduling so to relax given train schedule, linearize a train fuel consumption model, and integrate them to LAP model for train/locomotive energy optimization model (marked as LAP-Fuel).

1.3 State of Art

In this section, we survey the most significant works on LAP, the acceleration for CG decomposition, LAP with train re-scheduling and fuel consumption model. The full details of literature review are available in the corresponding literature review section of Chapter 2, 3 and 4.

1.3.1 Locomotive Assignment

For LAP model, Ahuja *et al.* [9] developed a mixed integer linear program (MILP) model of LAP in the for CSX Transportation, one of the major freight railway companies in US. They used similar objective and network structure as Ziarati *et al.* [55], considering most of the general constraints, but excluding the locomotive maintenance process. To solve the scalability issue, a neighborhood search heuristic was proposed for large scale scenarios. In addition, the consist busting issue was not considered by their model.

The model of Ahuja *et al.* [9] suffered from high rate of consist busting, so Vaidyanathan *et al.* [54] developed a consist based LAP model. The model used a pre-processing algorithm to generate locomotive consist configurations for the main model. Their model reduced consist busting rate but did not consider maintenance process. In addition, it still had potential scalability issue, i.e., larger data set scenarios would need much more consist configurations than the pre-processing algorithm can develop (which would be solved by CG decomposition technique, e.g., in our LAP model).

Cacchiani *et al.* [17] provided two integer linear programming (ILP) models for a similar problem: train unit assignment problem. The models combined cars and locomotive for certain trips, but without the maintenance constraint.

Cordeau *et al.* [23] proposed an exact model for locomotive and car assignment problem for passenger trains. The model, which was an extension of their previous work [22], covered all general constraints, e.g., maintenance constraint, and also considered the consist busting (they call it switching). This model has the potential to be converted to general locomotive assignment problem for freight trains, but firstly it has a limited number of train consist configurations as well as train sequences, secondly it has one maintenance center and assumes that during the scheduled time each equipment has to be maintained at least once. In freight trains the situation is more complex (details in Chapter 2).

For CG decomposition technique, there exist some strategies in different stages to accelerate computational time and/or the convergence rate. Desaulniers *et al.* [24] decomposed large problem into small subproblems, and merge the solutions after to reduce the usage of time and memory. Sadykov *et al.* [50] prepared a good-enough initial solution in problem pre-process stage. Chen *et al.* [19] pre-created the columns pool by problem-specific knowledge. There are schemes allow a pricing problem model to return multiple columns with negative reduced cost, e.g., Goffin *et al.* [30] gained efficiency from adding multiple cuts central cuts simultaneously.

The general LAP deals with given and fixed train schedules. Obviously if adjustment of train schedule is allowed, it will have more flexibility and potential efficiency. Fügenschuh *et al.* [26] establish a model for locomotive and car cycle scheduling problem with time window. It allows train delay within given time window but ignore locomotive maintenance. In addition, there is no explanation for how to determine the time length of window and why.

1.3.2 Train Energy Consumption

Train fuel consumption problem has been investigated for a long time, and different models are established (details in Section 4.5). Based on these previous works, we can calculate the fuel consumption of given train with parameters, e.g., weight, type, speed, distance, *etc.*. And the train speed is directly determined by the output power of the locomotives. So based on the discussion in Section 1.2.2 & 1.2.3, for a freight train, since other parameters are fixed by the schedule, its fuel consumption can be controlled by the adjustment of speed via the adjustment of locomotive power assigned and rescheduling. If a LAP model can cooperate with a train fuel consumption model, it is possible to optimize the total train energy consumption through locomotive assignment and speed control. But almost no research work has been done in this particular area.

1.4 Thesis Contributions

The contributions of this thesis have been published in the following papers, each in one chapter. Here below is a brief description:

Contribution #1 (published [34] & submitted [37]): we propose a consist travel plan based LAP model with CG decomposition. The model considers constraints needed in train/locomotive operation including locomotive maintenance constraint and consist busting avoidance. The key characteristics of our LAP model are that firstly the model uses the number of consist travel plans to evaluate consist busting occurrence, and secondly the flow conservation constraint can change locomotive flow status/type from critical to regular. The model can solve given 2-week train schedule including 1,750 trains in affordable time.

Contribution #2 (published [33] & submitted [52]): we establish a new Column Generation scheme to generate more than one consist travel plan configuration in each iteration. Each column, i.e., consist travel plan configuration is an optimal solution for a reduced sub-network that extracted in the pre-processing stage. Comparing to our previous LAP model, the new multi-CG LAP model saves 60-93% of computational time without reduce the quality of final solution. The new CG scheme has the potential to expand to some general CG formats of network flow problems.

Contribution #3 (submitted [53]): in this paper, we develop a collision-free algorithm for train schedule pre-processing, and corresponding constraints in LAP model, to relax train schedule in the calculated time windows for LAP model. Then the LAP model is integrated with a linearized trian/locomotive fuel consumption model to optimize the total locomotive energy consumption in given train schedule.

1.5 Organization of the Thesis

The thesis is organized into five chapters. Chapter 1 provides the basic background information of railway industry, the statement of the main research problem, some latest literature review and a brief list of our contributories. The next three chapters, each of which contains one major paper we published or submitted: Chapter 2 for LAP model with maintenance constraints and consist busting avoidance; Chapter 3 for our CG acceleration algorithm, i.e., Multi-CG architecture; Chapter 4 for the optimization of train/locomotive energy consumption on LAP and train scheduling level. Finally Chapter 5 draws the general conclusion of this PhD project and provides suggestions of future research works.

Chapter 2

A New Model for Locomotive Assignment under Consist Busting

B. Jaumard, H. Tian, and P. Finnie. submitted for publication, 2017. An extended abstract of this paper has been published in Joint Rail Conference (JRC), 2014 [34].

2.1 Introduction

Freight train transportation is highly investigated with high energy efficiency and safety advantage. This paper focuses on one of the key elements of railway operation problems: the locomotive assignment problem (LAP), which assign adequate power of locomotives to scheduled freight trains, with the proper constraints and rules. The LAP model optimizes the number and/or cost of locomotives to given train schedule while all the technical and business constraints are satisfied.

First, some basic terminologies, rules and constraints used in the description of LAP are recalled. A freight railway company usually has several types of locomotives with different attributes. A locomotive can be *active* (i.e., pulling a train), *deadheading* or *idling*. Deadheading locomotives operate under their own power but do not pull trains. Usually more than one locomotive are used to pull a train. Those locomotives which are grouped together to pull a train are called *consist*. In locomotive management, there is a time-consuming process called *consist busting*, i.e., disassembling the consist of an inbound train into stand alone locomotives and reassigning them to several outbound trains. Consist-busting means additional labor & operational cost and time requirements. It also reduces the robustness of the train schedule because it allows an outbound train to get locomotives from multiple inbound trains. If any of the source inbound trains is delayed, the outbound train has to be delayed as well. So consist busting should be avoided as much as possible.

Several trains, which are pulled by the same consist, build a *consist travel plan* without consist busting. The locomotives should undergo regular maintenance examinations in a shop station. The locomotive due to maintenance in the planning period under study is called *critical*. In order to follow the US legislation, a locomotive which misses its regular maintenance has to be turned off and towed to the shop. The maintenance rule is based on either calendar days or operation mileage. In this paper we take the time rule that the locomotive should be maintained in a shop every 90 days, which is the usual constraint in North America.

Several mathematical models and solution methodologies have been proposed for locomotive assignment. Most of them based on a multi-commodity network flow formulation and are either locomotive-based or consist-based. Both compact MILP (mixed integer linear program) or column generation formulations were investigated for exact solutions. Some other research studies investigated heuristics and approximate approaches to overcome the computational complexities. In the next section the details of those solution algorithms are provided.

After the literature review section, the paper recalls the statement of LAP, and provides the details of our consist travel plan base model. We next present the solution scheme of the proposed model, i.e., column generation decomposition algorithm that includes a consist travel plan generator, and consequently allows the consideration of all possible consist travel plans through an *implicit* enumeration. Numerical results are based on CPR's data sets spanning several time periods over the entire CPR network, from Vancouver to Montreal, including the U.S. railway component. Conclusions are drawn in the last section.

2.2 Literature Review

2.2.1 LAP with Exact Model

Ziarati *et al.* [55], [56]

Ziarati *et al.* [55] focuses on LAP with different types of locomotives. The LAP problem is reformulated as an integer multi-commodity time-space network flow problem, with a Column Generation (CG) decomposition. Each node is associated with a railway station and a particular time, the arcs represent activities such as waiting periods, train travel between two stations (usually the origin and the destination) or train maintenance, and commodities are the locomotives. The model comes with the general constraints including horsepower/tonnage, locomotive remotely transferring and maintenance constraints, but without the consideration of consist busting process and its cost. The model also allows the outpost process (which is the duty for a locomotive in a branch of railway that perform local pick-up and delivery process). The objective is to minimize the total cost of the locomotive assignment to the trains, i.e., the locomotive schedule.

The test infrastructure contains 26 stations, 164 outposts and 18 shops, which comes from Canadian National North America. The model can handle 1,249 locomotives of 26 types including 171 critical ones, and satisfies 1,988 train-segment requests, 238 outpost requests and 56 shop requests. The locomotive assignment schedule is non-periodic, solved as a set of overlapped exact problem as 2 or 3-day length, and finally merged to a 7-day horizon.

Ziarati *et al.* [55] improve the previous model by a heuristic branch first, cut second approach, which reduces 1.1% of locomotive usage and more than 20% of gap.

Rouillon *et al.* [48]

The authors extend the solution algorithm of Ziarati *et al.* [55] and [56] with three different branching methods and search strategies to develop a branch-and-price algorithm for LAP of a freight railway on operational level. The new extension of previous models saves 10 and 39 locomotives.

Ahuja *et al.* [9]

The authors develop a MILP for LAP in the planning level of CSX Transportation. They also formulate LAP as a locomotive flow model, i.e., an integer multi-commodity flow problem. The objective and network architecture are similar to the one of Ziarati *et al.* [55]. The differences are:

- The train schedule and the locomotive assignment plan is assumed to be cyclic every week, which means that the distribution of locomotives at the end of the schedule time period should be back to its distribution at the beginning of the planning period.
- Locomotive light travels (locomotive(s) relocation without being attached to a train) are allowed. This makes the model more general and flexible.
- The locomotive maintenance process is not considered.

Instead of using CG reformulation to solve the MILP model, the authors develop a neighborhood search algorithm/heuristic to improve the performance for large scale data instances.

The model has been validated with CSX data files. Each test scenario has 3,324 trains, 119 stations and 3,316 locomotives, 5 types of locomotives. A heuristic solution was produced within 30 minutes. The result use at least 40% fewer locomotives than the solution from CSX current software.

Vaidyanathan *et al.* [54]

To resolve the issues of these previous locomotive based models, i.e. the high rate of consist busting (the medium consist busting rate is 50%) and the severe scalability issues, Vaidyanathan *et al.* [54] extend the model of Ahuja *et al.* [9]. To minimize consist busting, Vaidyanathan *et al.* [54] focus on a consist-based assignment model. The configurations of consists are generated by a pre-processing algorithm. Then, instead of assigning locomotives, the model of Vaidyanathan *et al.* [54] assigns consists to pull the scheduled trains with respect to the minimum power and other business constraints.

There are two sets of constraints in the consist-based formulation. The first set of constraints correspond to the so-called hard constraints, i.e., mandatory constraints in order to reach a feasible solution, e.g. consist length limit, consist disjoint constraints, *et al.*. The soft constraints focus on the avoidance of consist busting, of single locomotive consist, and of train/consist configuration rebuilding on different days (i.e., always assigning the same consist to the same train). Other general requirements of the normal LAP are implicit, hidden in the configuration of consists: those guarantee that the requirements of each train for the horsepower, tonnage, limit of active axle, consist size, and of locomotive type for train & terrain.

Their consist-based formulation uses a data set with 382/388 trains, 6 locomotive types, 87 stations, and 3 up to 17 types of consists in the test scenarios. The computational time for each test scenario is up to 450 seconds on a Pentium IV desktop computer. The quick convergence rate for consist-based model has several reasons. The first reason is that the consist-based formulation applies some hard constraints implicitly, e.g., the requirements of tonnage, power and the limit of 24-active axle. The second reason is that the number of decision variables is smaller than that of locomotive-based models. The last reason is that the decision variables for active consist assignment are binary, instead of the general integer variables in locomotive-based formulation.

The potential issue of the consist-based formulation is: the greater optimization of the solution requires greater numbers of configuration types of consist. However, the computational time will grow as well, and even faster (it is true for normal MILP, but can be solved by CG decomposition as we proposed in next chapter). The other issue is that Vaidyanathan *et al.* [54] do not consider the maintance/shopping constraints for locomotives, and only assume the locomotive plan is cyclic every week. In addition, their consist-based model does not take the initial locomotive positioning into account. Instead, the authors propose a post-processing algorithm which minimizes the locomotive repositioning (light trains) in order to ensure a smooth transition from the current locomotive locations to the locomotive location requirements in the solution of the LAP model. This entails a lot of locomotive reposition issues.

Piu [44]

Piu [44] considers the fueling and maintenance constraints as well as the robustness, in order to make the final LAP solution easier. A preliminary optimization model is prompted to initialize the types of consist for the final LAP model. But unfortunately no numerical results are presented, only a synthesis of the data sets and of the numerical results of the recent papers.

Cordeau *et al.* [22], [23]

Cordeau *et al.* [22] establish a MIP model for locomotive and car assignment problem, which covers the general constraints except the maintenance constraints. The model allows deadheading for locomotive relocation, uses a given set of consist types, but does not consider the extra cost of consist busting. Benders decomposition technique is applied to solve the MIP model with large scale data set. The largest test case from VIA Rail comes with 348 train legs, 32,981 sequence variables, 105,327 arc variables and 205,080 constraints.

In [23], the authors extend the previous model, with the consideration of locomotive maintenance, and the consist busting, which is called switching. A two-phase method is proposed, in which based on the solution of general model in phase I, an optimization for the consist switching is processed in phase II. A branch-and-bound solution method with CG is applied for the model. Six instances of more than 300 VIA Rail passenger trains in the Quebec-Windsor corridor are solved with good quality solutions within several hours.

Maposa and Swene [41]

Maposa *et al.* [41] solved the planning locomotive scheduling model (LSM) of National Railways of Zimbabwe (NRZ) with multiple locomotive types. They regarded LSM as an integer multi-commodity flow problem that is similar to that of Ahuja [9]. But all (freight) trains of NRZ started and ended at the station Mpopoma, since it was the only marshalling yard. Their mixed integer programming models considered the deadheading, light traveling and consist busting conditions, and can solved 528 trains (in 3 different types)with 2 locomotive types.

2.2.2 LAP with Heuristic

Godwin *et al.* [29]

The authors study the effects on locomotive assignment by locomotive fleet size for freight trains in Indian railway networks. Their environment is special as the freight trains have day-to-day schedule, and have lower priority than the passenger trains in the shared railway network. The authors develop first a Petri-Net model then a heuristic based on it for locomotive assignment including deadheading. They test the model in a five-division railway network with 467 stations (78 of them are loading/unload stations) in which 127 trains per day generate locomotive requests. The time duration is 50 days. Their conclusion is that the efficiency of locomotive assignment is greatly affected by the fleet size and the holding time of locomotives before deadheading permission.

Ghoseiri *et al.* [27]

The authors focus on the homogeneous locomotive assignment problem with deadheading and maintenance processes. They formulate the problem into the well known vehicle routing problem with time windows (VRPTW). To solve VRPTW with the initial status that the locomotives are distributed into each depot/station, they develop a cluster-first, route-second approach. First, they decompose the original multi-depot locomotive assignment into a series of single depot problems (A depot is a home station of locomotives that serve the neighbor stations). And then each of such problems is independently solved by a hybrid genetic algorithm, i.e., using a push forward insertion heuristic to create initial solutions and then applying the genetic algorithm to improve the solutions. An artificial, randomly generated test frame of 84 stations and 42 trains per day is used to test their algorithm over a up-to-7-day time horizon. The test frame has 10 locomotives in only 1 depot. The solutions of different number of trains shows the empirical complexity of the algorithm is between $O(n^2)$ and $O(n \log n)$. For the solution quality of their algorithm, they compare it with a branch & bound algorithm for a few small and medium sized problems. The results show their algorithm has zero error and much less computational time.

Powell *et al.* [46]

The authors develop the Princeton Locomotive and Shop Management system for Norfolk Southern Corporation. They point out the scalability issue of the deterministic optimization models for LAP, especially for long-term locomotive planning, and develop a framework of an approximate dynamic programming (ADP) approach, for short to long-term LAP. Their paper considers the shop routing problem, transit time delays, dynamic schedule changes, and equipment failures.

Using ADP, the authors decompose the large scale deterministic LAP over time. Each subproblem is to assign locomotives to trains on a short time horizon (e.g., 4-6 hours), and is formulated as an integer linear problem. The authors attempt to solve each subproblem exactly and also to consider the impact of solution of each subproblem on the future. The key idea is that although the exact impact of the solution of subproblem on the future can not be computed, the approximate value could be learned and calibrated by running the model iteratively comparing a truly realistic model. The calibration of the model of Powell *et al.* [46] has been processed during several years over historical data. Now their system works properly by Norfolk Southern for fleet sizing studies. However, as soon as the traffic will change, calibration may need to be updated.

Noori *et al.* [43]

Noori *et al.* [43] solve LAP with homogeneous locomotives in several depots, and with pre-scheduled trains of different priorities. They develop a two-phase approach in which they decomposed the main problem into multiple single depot subproblems, and solve each subproblem by a hybrid genetic search algorithm heuristically, with the consideration of locomotive maintenance. Part of the (medium size) results were compared with branch-and-bound algorithm to prove its correctness and efficiency. The maximum size of the LAP problem they solved is 280 trains within one week in a 80 nodes network.

Kasalica *et al.* [38]

Kasalica *et al.* [38] allow train delays in the cyclic locomotive assignment on the Serbian Railways and Montenegrin Railways networks. They propose a 5-stage algorithm/heuristic, without the maintenance constraint, in order to minimize the total locomotive process and idle time in the station. Their algorithm are tested by 72 trains with 13 locomotives.

2.2.3 Related Problem

Cacchiani *et al.* [16], [18]

Cacchiani *et al.* [16] focus on the train-unit (TU) assignment problem (TUAP) that to assign a set of TU to satisfy trips of seats demands. A TU is a self-contained train with an engine and passenger seats. They allow TU to light travel and consider the maintenance. But the time and cost to dassemble/reassemble TUs from one trip to another are not considered. A LP-based heuristic method is proposed to solve the problem. The method is tested by up to 660 trips with 10 TU types.

In [18], they propose a Lagrangian heuristic algorithm to solve TUAP, and can solve larger realistic instances up to 1190 trips with 18 TU types.

2.3 Problem Statement

The present study provides a consist travel plan based LAP optimization model, with the technical, industrial and business constraints mentioned above, including maintanance/shop constants and consist busting avoidance. In addition, in order to guarantee that the model is always able to output a solution, we allow more locomotives out of given fleet and penalize the extra requirement.

The proposed model, which will be detailed in the next section, builds an assignment of locomotives with respect to the given constraints. The input of LAP model is divided into two parts: train schedules and parameters, and locomotive parameters. Train schedules include the information of the departure/arrival times and the origin/destination stations of each scheduled train. Characteristics of the trains such as power requirement are also given. Locomotive parameters include their horsepower, their initial locations and types. Maintenance shop capacities are given, i.e., the number of locomotives that can undergo maintenance in each shop per day. We assume that the train schedule is fixed: the arrival/departure time in each station is known and cannot be modified. We consider all possible consists (throughout an implicit enumeration thanks to the column generation techniques).

The output of the model is the travel plan for each locomotive within a series of consist assignments, as well as the routes of deadheading (or light traveling) locomotives that connect the consecutive routes within an active consist in each train.

2.4 Locomotive Assignment Model

The proposed LAP model relies on a multi-commodity network, which we next describe, before detailing the variables, the objective and the constraints of the consist-based model.

2.4.1 Multi Commodity Network

The multi commodity network is a time/space network, see Figure 1, where each node v is associated with two components: $\textsc{location}(v)$, which corresponds to a railway station location, and, $\textsc{time}(v) \in \mathrm{Z}^+$, which corresponds to the beginning or the end of an activity, and which is expressed in minutes. The arcs represent activities such as waiting periods, train travel between two stations (usually the origin and the destination) or train maintenance, and commodities are the locomotives. It is sometimes convenient to represent a multi-commodity network as a layered graph, where each layer is associated with one commodity, as some nodes and arcs may be specific to a particular commodity. We will assume here the same network for all types of locomotives.

In order to built the multi-commodity network for a given planning period, one has to take into account some legacy trains, i.e., trains that departed in the previous planning period but which arrive in the current planning period. Similarly, there are trains departing in the planning period under study, which will reach their destinations after the end of the planning period under study. For instance, in Figure 2, Train 1 departs before the beginning of the planning period under study.

We now describe in detail the generic multi-commodity network $G = (V, L)$ associated with the overall set of locomotives. V denotes the set of nodes, indexed by v, where each v has a space and a time coordinate. L is the set (indexed by ℓ) where

$$L = L^T \cup L^{\textsc{shop}} \cup L^W \cup L^D$$

with the subsets defined below.

- L^T is the set of links associated with trains.

- $L^{\textsc{shop}}$ is the set of links associated with the three month maintenance activity (regulatory requirement in North America), which takes place in a station with a yard. For $\ell \in L^{\textsc{shop}}$, its time origin is always 8am, and its time destination is 5pm, so that it corresponds to one working day. In other words, $\ell = (\textsc{location}, 8 \text{ am}) \rightarrow (\textsc{location}, 5 \text{ pm})$ for $\textsc{location}$ being a shop with a yard.

- L^{W} is the set of locomotive waiting links, i.e., associated with idle periods for a locomotive: $\ell = (v, v')$, where nodes v and v' are associated with the same station, i.e., $\textsc{location}(v) = \textsc{location}(v')$, if locomotive κ is idle in station $\textsc{location}(v)$ between $\textsc{time}(v)$ and $\textsc{time}(v')$. In addition, L^{W} includes the set of virtual links that are pointing to the dummy sink node v^{SINK}.

- L^{D} is the set of locomotive deadheading links, i.e., the same time/space parameter of the corresponding train link but the different activity: a critical locomotive can go through the deadheading link after the due date.

Among the nodes, we identify the so-called source and destination nodes as follows:

V^{SRC}: indexed by v^{SRC}, as the set of nodes (stations in the railway system) where some locomotives are first available in the planning period. It is either a node with time component equal to the beginning of the planning period if a locomotive is idle in the corresponding station, or at the departure time of a (legacy) train prior to the beginning of the planning period, with the arrival time after the beginning of the scheduling time period, see Figure 2. Therein, V^{SRC} contains 6 nodes, 5 with a time index equal to the origin of the planning period, and one at an earlier time.

v^{SINK}: dummy destination node, where all destination arcs converge. See the links represented by the long dashed lines in Figure 1 for an illustration.

The set of nodes contains all the endpoints of the links of L. For each $\ell = (v, v') \in L^{\text{SHOP}}$, there is another link $\ell' = (v, v') \in L^{W}$ in order for a locomotive to bypass the maintenance step if not required or if the shop is full.

2.4.2 Notations

S is the set of consist travel plans, where a consist travel plan $s \in S$ defines a sequence of trains led by the same locomotive consist.

$$S = \bigcup_{v \in V} S_v^+,$$

where S_v^+ denotes the set of consist travel plans originating at v. Similarly, S_v^- denotes the set of consist travel plans destined to v. K is the set of locomotive types, indexed by k. When we need to distinguish critical from regular locomotives, we use the index k_c (resp. k_r), and decompose K into K_c and K_r. Critical locomotives are those ones whose operation days will reach the calendar maintenance time limit during the planning time period.

We denote by n_k the number of available locomotives of type k throughout the network, and

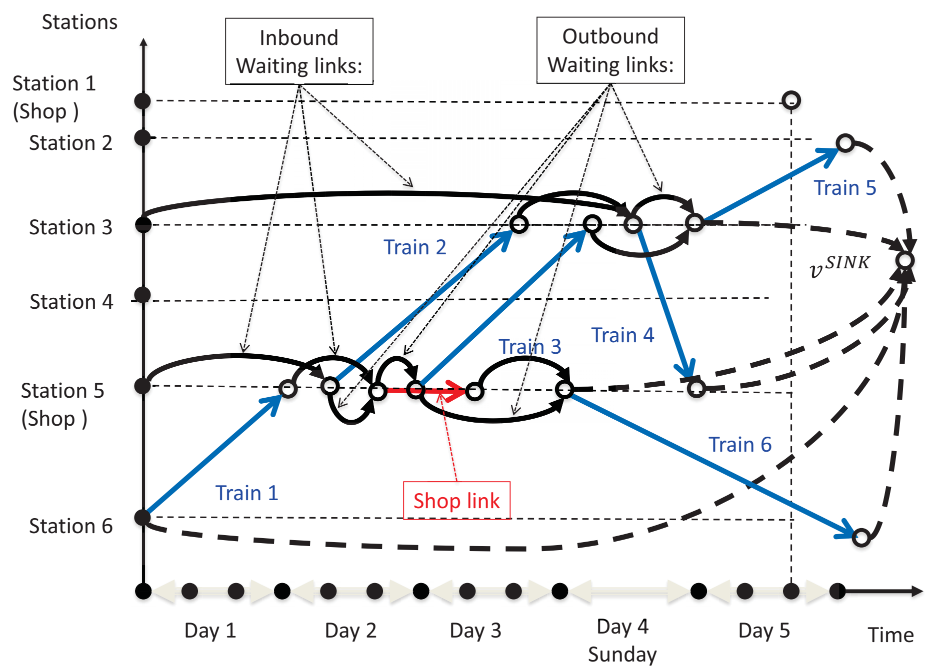


Figure 1: Multi Commodity Network

n_k^s the number of locomotives of type k of each consist travel plan $s \in S$, and
$n_{k,v}^{\text{SPARE}}$ the spare locomotives of type k in each start node $v \in V^{\text{SRC}}$.
$\text{DST}_v^s = 1$ if consist travel plan s ends at node v (final endpoint of a train link), 0 otherwise
$d_\ell^s = 1$ if train link $\ell \in L^T$ belongs to consist travel plan s, 0 otherwise
$n_{k,v}^{\text{SPARE}}$ = number of spare locomotives of type k in source node $v \in V^{\text{SRC}}$
$\text{CAP}(\ell^{\text{SHOP}})$ = upper bound of critical locomotives that can be maintained in shop link $\ell^{\text{SHOP}} \in L^{\text{SHOP}}$.
$\text{TimeSrc}(t), \text{TimeDst}(t)$ = the start and end time of train t, counted from the start time of LAP scheduling period.
m_k = total number of calendar days for the locomotives of critical type $k \in K_c$ since their last visit to a shop, until the start time of LAP scheduling period. In order to alleviate the description of the model, we assume that all critical locomotives of a given type have the same value m_k. However, if it is not the case, it is easy to expand the model, by defining as many sets K_c as the number of different values for the total number of calendar days since the last visit to a shop for all critical locomotives.

In the multi-commodity graph $G = (V, L)$, we designate by $\omega(v)$ (resp. $\omega(V')$ with $V' \subseteq V$) the set of adjacent links to v (resp. to a node of V'). In addition, $\omega^+(v)$ (resp. $\omega^-(v)$) denotes the set of adjacent outgoing (resp. incoming) links of v. For a given link ℓ,

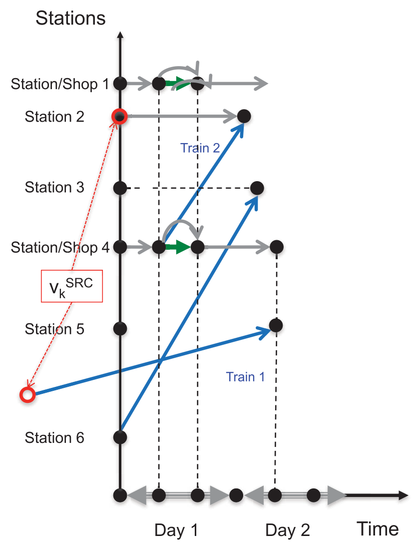


Figure 2: Definition of the V^{SRC} Source Node Set

$\delta^+(\ell)$ denotes the destination endpoint of ℓ, and $\delta^+(L'), L'' \subseteq L$ denotes the set of destination endpoints of the links of L'. Similarly, $\delta^-(\ell)$ and $\delta^-(L'), L'' \subseteq L$ denote the origin endpoint(s) of ℓ and of the links of L', respectively.

2.4.3 Consist Travel Plans and Waiting Links

A consist travel plan is defined as a set of trains that use the same locomotive consist one train after the other one, without any consist busting. The number of consist busting therefore can be evaluated by consist travel plans occurrences, for that the end of a consist travel plan means a breakup of consist. The idea is similar to Cordeau *et al.* [23], which use train sequence and switching in passenger trains.

A train link can belong to at most one plan. If a train does not belong to any plan, then it means there is a lack of available locomotives in order to pull it. Consist travel plans (and shop links) are separated by waiting links, and must be spaced a minimum time (2 hours in our numerical experiments) in order to allow consists to be busted and reassembled. Within a consist travel plan, the time difference of two consecutive train links is spaced by a minimum time period, at least 1 hour in our numerical experiments.

To ensure this time constraints between, e.g., two successive train departures, or for reassembling a consist, we divide the waiting links into the inbound and the outbound waiting links: $L^{\text{W}} = L^{\text{W\_IN}} \cup L^{\text{W\_OUT}}$ with $L^{\text{W\_IN}} \cap L^{\text{W\_OUT}} = \emptyset$. In addition, such a division will allow us to identify the consist busting.

A waiting link is defined as a link with its two endpoints associated with the same location (station) components. An ***inbound waiting link*** ($\ell^w \in L^{\text{W\_IN}}$) starts at the destination node of a train link or at a source node, and ends at the nearest origin node of another train/shop link, and no less than certain time interval. Since there are two different time interval requests for inbound waiting links: the smaller one represents a time duration that is at least the time required to re-assign a given consist to another train, and the larger one represents the minimum time to re-assign a locomotive to another consist, for each destination node of a train link, we have create two inbound waiting links: one to the nearest origin node of another train/shop link no less than the smaller interval request, the other to satisfy the larger time interval request. An ***outbound waiting link*** **(**$\ell^w \in L^{\textbf{w\_out}}$**)** starts and ends at the nearest origin nodes of two different train links, or at an origin node and the dummy sink node, without any time restriction. Figure 1 illustrates the division of waiting links. Shop links are considered as train links, with respect to the definition of inbound/outbound waiting links.

2.4.4 Variables.

We use four sets of variables:
$z_s = 1$ if consist travel plan s is selected, 0 otherwise.
x_{kv}^{NEED} = number of additional required locomotives of type k at source node $v \in V^{\text{SRC}}$ in order to be able to assign adequate locomotives to all trains.
$x_{k\ell}^{\text{LOCO}}$ = number of locomotives of type k on link ℓ. Note that:

$$x_{kL^{\text{W}}}^{\text{LOCO}} = x_{k_r L^{\text{W}}}^{\text{LOCO}} + x_{k_c L^{\text{W}}}^{\text{LOCO}}, x_{k\ell^D}^{\text{LOCO}} = x_{k_r \ell^D}^{\text{LOCO}} + x_{k_c \ell^D}^{\text{LOCO}},$$

$$x_{k_r \ell^{\text{SHOP}}}^{\text{LOCO}} = 0, \quad x_{k\ell^{\text{SHOP}}}^{\text{LOCO}} = x_{k_c \ell^{\text{SHOP}}}^{\text{LOCO}}.$$

2.4.5 Maintenance and Shops

A critical locomotive must stop at a shop for maintenance operations every maintenance period, here assumed to be a calendar interval, see, e.g., Railway Locomotive Inspection and Safety Rules [5] for more details on maintenance intervals. Critical locomotives are relabeled as regular after completing the maintenance process at a shop. This relabeling will be taken care thanks to special flow conservation constraints at the shop end nodes in the proposed LAP model.

2.4.6 Optimization Model

The LAP optimization model we propose is under the assumption without the legacy trains.

2.4.7 Objective

The objective is to minimize the size of the locomotive fleet required as well as to reduce consist busting, although these two segments seem in the opposite directions. Furthermore, the consist busting can not be counted implicitly, but since each consist busting brings at least two additional consist, the number of consist travel plans increase correspondingly. We therefore propose the following objective with the minimization of: *(i)* the number of consist travel plans with weight (PENAL_z); *(ii)* the number of total locomotives in operation; and *(iii)* the number of extra locomotives (details in Section 2.4.8) with penalty (PENAL_k).

$$\min \quad \sum_{s \in S} \text{PENAL}_z \cdot z_s + \sum_{\ell \in \omega^-(v^{\text{SINK}})} \sum_{k \in K} x_{k\ell}^{\text{LOCO}} + \sum_{v \in V^{\text{SRC}}} \sum_{k \in K} \text{PENAL}_k \cdot n_{k,v}^{x_{kv}^{\text{NEED}}} \tag{1}$$

2.4.8 Constraints

Flow Conservation Constraints. In any normal node in the locomotive flow, there should be equal number for outbound locomotives and inbound ones for each type:

$$\begin{aligned} &\sum_{s \in S_v^+} n_k^s \, z_s + \sum_{\ell \in \omega^+(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x_{k\ell}^{\text{LOCO}} \\ &\qquad = \sum_{s \in S_v^-} n_k^s \, z_s + \sum_{\ell^w \in \omega^-(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x_{k\ell}^{\text{LOCO}} \\ &\qquad\qquad v \in V \setminus \left(V^{\text{SRC}} \cup v^{\text{SINK}} \cup \delta^+(L^{\text{SHOP}})\right), k \in K_r \cup K_c \end{aligned} \tag{2}$$

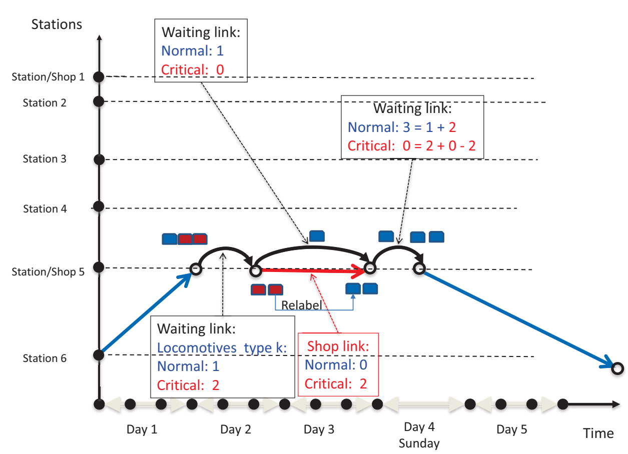


Figure 3: Flow Conservation of Locomotive Type k for Shop Link

The key idea of our proposed LAP model is the critical locomotive **re-labeling** process, i.e., at the destination node of a shop link, the critical locomotives after maintenance should be re-labeled as regular ones, of the same type, as shown in Figure 3:

$$
\begin{aligned}
&\sum_{s \in S_v^+} n_{k_r}^s\, z_s + \sum_{\ell \in \omega^+(v) \cap (L^{\text{WAIT}} \cup L^D)} x_{k_r \ell}^{\text{LOCO}} \\
&\quad = \sum_{s \in S_v^-} n_{k_r}^s\, z_s + \sum_{\ell \in \omega^-(v) \cap L^{\text{SHOP}}} x_{k_c \ell}^{\text{LOCO}} + \sum_{\ell \in \omega^-(v) \cap (L^{\text{WAIT}} \cup L^D)} x_{k_r \ell}^{\text{LOCO}} \\
&\qquad\qquad v \in \delta^+(L^{\text{SHOP}}), k = \{k_r, k_c\} \in K
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
&\sum_{s \in S_v^+} n_k^s z_s + \sum_{\ell \in \omega^+(v) \cap (L^{\text{WAIT}} \cup L^D)} x_{k\ell}^{\text{LOCO}} = \sum_{s \in S_v^-} n_k^s z_s + \sum_{\ell \in \omega^-(v) \cap (L^{\text{WAIT}} \cup L^D)} x_{k\ell}^{\text{LOCO}} \\
&\qquad\qquad v \in \delta^+(L^{\text{SHOP}}), k \in K_c.
\end{aligned} \tag{4}
$$

In addition, the flow conservation constraints above allows that two consist travel plans are connected directly to reuse (part of) the same locomotives. Constraints (5), (6), (7) guarantee to avoid this issue, i.e., in any node v that sent out a consist travel plan s, the locomotives that v assigns to s can only be those from the waiting links or shop links that end in v, the locomotives brought by the consist travel plans that end in v are not eligible.

$$
\begin{aligned}
&\sum_{s \in S_v^+} n_k^s\, z_s \le \sum_{\ell^w \in \omega^-(v) \cap L^{\text{WAIT}}} x_{k\ell}^{\text{LOCO}} \\
&\qquad v \in V \setminus \left(V^{\text{SRC}} \cup v^{\text{SINK}} \cup \delta^+(L^{\text{SHOP}})\right), k \in K_r \cup K_c
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
&\sum_{s \in S_v^+} n_{k_r}^s\, z_s \le \sum_{\ell \in \omega^-(v) \cap L^{\text{SHOP}}} x_{k_c \ell}^{\text{LOCO}} + \sum_{\ell \in \omega^-(v) \cap L^{\text{WAIT}}} x_{k_r \ell}^{\text{LOCO}} \\
&\qquad v \in \delta^+(L^{\text{SHOP}}), k = \{k_r, k_c\} \in K
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
&\sum_{s \in S_v^+} n_k^s z_s \le \sum_{\ell \in \omega^-(v) \cap L^{\text{WAIT}}} x_{k\ell}^{\text{LOCO}} \\
&\qquad v \in \delta^+(L^{\text{SHOP}}), k \in K_c.
\end{aligned} \tag{7}
$$

Spare Locomotive Constraints. The LAP model may get infeasible solution, i.e., in any station, the total locomotives required by the solution exceed the given spare locomotives at the start of the planning period. In order to avoid this situation, we introduce the variable x_{kv}^{NEED} that count the number of extra locomotives in order to accommodate all scheduled trains, and the flowing constraints (The last set of constraints (10) guarantee that even we allow the additional locomotive in use, the total locomotives in operation still can not exceed the maximum number of locomotives in each type. The numbers used in test scenarios are displayed in Table 1):

$$\sum_{s\in S_v^+} n_k^s\, z_s + \sum_{\ell^w\in\omega^+(v)} x_{kL^{\mathrm{W}}}^{\mathrm{LOCO}} + \sum_{\ell^w\in\omega^+(v)} x_{k\ell^D}^{\mathrm{LOCO}} - x_{kv}^{\mathrm{NEED}} \le n_{k,v}^{\mathrm{SPARE}} \qquad k\in K_r, v\in V^{\mathrm{SRC}} \tag{8}$$

$$\sum_{s\in S_v^+} n_k^s\, z_s + \sum_{\ell^w\in\omega^+(v)} x_{kL^{\mathrm{W}}}^{\mathrm{LOCO}} + \sum_{\ell^w\in\omega^+(v)} x_{k\ell^D}^{\mathrm{LOCO}} \le n_{k,v}^{\mathrm{SPARE}} \qquad k\in K_c, v\in V^{\mathrm{SRC}} \tag{9}$$

$$\sum_{\ell\in\omega^-(v^{\mathrm{SINK}})} x_{k\ell}^{\mathrm{LOCO}} \le n_k \qquad k\in K \tag{10}$$

Train-Disjoint String Constraints. Each train should belong to **exactly** one consist travel plan :

$$\sum_{s\in S} d_\ell^s\cdot z_s = 1 \qquad \ell\in L^T \tag{11}$$

Shop Capacity Constraints. For each station with shop, there is a limit for critical locomotives allowed at the same time for maintenance:

$$\sum_{k_c\in K} x_{k_c\ell^{\mathrm{SHOP}}}^{\mathrm{LOCO}} \le \mathrm{CAP}(\ell^{\mathrm{SHOP}}) \qquad \ell^{\mathrm{SHOP}}\in L^{\mathrm{SHOP}} \tag{12}$$

Consist-Busting Constraints. For any two consecutive consist travel plans , there is a minimum dwell time of at least $\mathsf{dwell\_loco}$, for the time required to break and re-assemble locomotive consists.

$$\sum_{k\in K} x_{kL^{\mathrm{W}}}^{\mathrm{LOCO}} = 0 \qquad \ell^w\in L^{\mathrm{W\_IN}}\setminus\omega^+(V^{\mathrm{SRC}}) : \mathsf{time}(\ell^w) < \mathsf{dwell\_loco}. \tag{13}$$

Deadheading The identification of the deadheading occurrences is done in post-processing phase, where we identify the deadheading locomotives which are not needed for pulling a specific train in consist travel plan but should be transferred to remote station to pull anther train. The data will be analyzed in Section 2.6.5.

2.5 Solution of the Locomotive Assignment Model

To solve the model developed in the previous section, we need to use column generation techniques in order to avoid the exhaustive enumeration of the consist travel plans , and limit their generation to the improving ones. We therefore use a solution scheme that is summarized in the flowchart of Figure 4, and which is described in next section.

2.5.1 Solution Process

We establish a first model with an initial set of consist travel plans , called Restricted Master Problem, while the Master Problem corresponds to the model described in the previous section. The Restricted Master Problem is solved alternately with a consist travel plan generator. Indeed, the consist travel plan generator corresponds to the so-called pricing problem in optimization, see, e.g., Chvátal *et al.* [20], which either generates an improving consist travel plan, i.e., a consist travel plan whose addition improves the value of the linear relaxation of the current restricted master problem, or concludes that the current solution of the RMP is indeed the optimal solution of the linear relaxation of the Master Problem. It then remains to generate an integer solution, which can be easily done using an iterative rounding off procedure. We next discuss how to define a generator of a consist travel plan such that its addition to the incumbent set of consist travel plans guarantee an improvement of the incumbent value of the linear relaxation of the model developed in the previous section. We first define the set of variables, then the objective of the consist travel plan generator, and then its set of constraints. We end this section with a flowchart Figure 4 summarizing the solution process.

2.5.2 Column Generation and Integer Solution

As illustrated in the flowchart of Figure 4, the solution process consists of two phases: the optimal solution of the linear relaxation of the master problem with the column generation technique, and then the derivation of an integer solution thanks to a branch-and-bound algorithm (CPLEX MILP solver) applied to the constraint matrix associated with the optimal solution of the linear relaxation of the master problem. While we are not guaranteed to generate a strict optimal solution, it allows the generation of an ε-optimal solution, where $\varepsilon = \frac{\tilde{z}_{i\bar{l}p} - z^\star_{\text{LP}}}{z^\star_{\text{LP}}}$, with $z^\star_{\text{LP}}$ being the optimal value of the linear relaxation of the Master Problem as described in (1)-(12).

The master problem associated with only a subset of the possible consist travel plans is called the restricted master problem. $\tilde{z}_{\text{ILP}}$ is the optimal integer solution of the last restricted

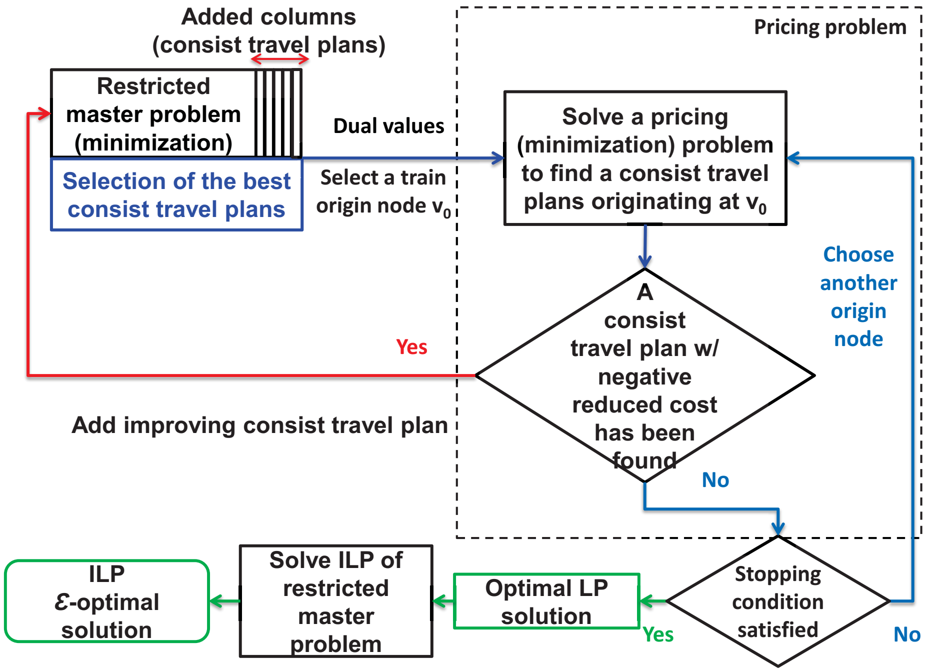


Figure 4: Flow Chart: Column Generation Process

master problem, i.e., the one with all the z_s variables generated until we reach the optimal solution of the linear relaxation of (1)-(12).

Note that the master problem is not solved optimally by embedding all possible consist travel plans , but only a very small subset of them. At each iteration of the column generation algorithm, one more consist travel plan is added, for a given node $v \in V$. Indeed, the column generation algorithm consists in a set of rounds, where, in each round, the algorithm goes through each source node of a train link, and check whether a consist travel plan with a negative reduced cost can be generated thanks to the pricing problem. If, during a round, the algorithm fails to find at least one consist travel plan with a negative reduced cost, the algorithm has reached the optimal solution of the linear programming relaxation, $z^{\star}_{\text{LP}}$.

2.5.3 Pricing Problem: Consist Travel Plan Generator

Consist travel plans are generated by the so-called pricing problem for a given origin node v_{O} for the consist travel plan under construction. In order to generate an exact solution, the column generation algorithm must consider all possible origin nodes, i.e., all nodes associated with the departure time of a train.

Variables

A consist travel plan starts with a train link, and ends at the endpoint of (another or the

same) train link.
$\text{DST}_v = 1$ if consist travel plan s under construction ends at node v, 0 otherwise, for $v \in \delta^+(L^T)$.
$x_\ell = 1$ if the link $\ell \in L$ belongs to the path supporting the consist travel plan, 0 otherwise. Note it is in one to one correspondence with the parameter d_ℓ of the master problem
n_k = number of locomotives of type k in the consist travel plan under construction
$n_k^d = 1$ if locomotives of type k is selected for the consist travel plan under construction, 0 otherwise
$y_{k,v} = n_k \cdot \text{DST}_v$. It is used to linearize the product $n_k \cdot \text{DST}_v$ of decision variables.

Objective: Reduced Cost of the z_s Variables

The objective of the pricing problem is the so-called reduced cost (if not familiar with linear programming concepts, the reader is referred to, e.g., Chvátal *et al.* [20]) of the z_s variables. The s index is omitted in this in order to alleviate the notations.:

$$\begin{aligned}
\overline{\text{COST}} = \sum_{k \in K} n_k \quad + \quad & \left[\sum_{k \in K_r} u_{kv}^{(8)} n_k + \sum_{k \in K_c} u_{kv}^{(9)} n_k \right] \\
& - \sum_{k \in K_r \cup k_c} \left(u_{kv_{\text{O}}}^{(2)} n_k - \sum_{v \in V \setminus (V^{\text{SRC}} \cup \{v^{\text{SINK}}\} \cup \delta^+(L^{\text{SHOP}}))} u_{kv}^{(2)} \cdot y_{k,v} \right) \\
& - \sum_{k \in K_r} \left(u_{k,v_{\text{O}}}^{(3)} \cdot n_k - \sum_{v \in \delta^+(L^{\text{SHOP}})} u_{k,v}^{(3)} \cdot y_{k,v} \right) \\
& - \sum_{k \in K_c} \left(u_{k,v_{\text{O}}}^{(4)} \cdot n_k - \sum_{v \in \delta^+(L^{\text{SHOP}})} u_{k,v}^{(4)} \cdot y_{k,v} \right) \\
& + \sum_{k \in K_r \cup k_c} \sum_{v \in V \setminus (V^{\text{SRC}} \cup \{v^{\text{SINK}}\} \cup \delta^+(L^{\text{SHOP}}))} u_{kv}^{(5)} \cdot y_{k,v} \\
& + \sum_{v \in \delta^+(L^{\text{SHOP}})} u_{k,v}^{(6)} \cdot y_{k,v} + \sum_{v \in \delta^+(L^{\text{SHOP}})} u_{k,v}^{(7)} \cdot y_{k,v} - \sum_{\ell \in L^T} u_t^{(11)} \cdot x_\ell. \qquad (14)
\end{aligned}$$

Note that when the pricing problem is solved for an origin node v_{O} that does not belong to VSRC, the term between square bracket in the above expression should be omitted.

Constraints of the Train String Generator

Ensure that there is exactly one destination node and it has to be the endpoint of a train link:

$$\sum_{v \in \delta^+(L^T)} \mathrm{DST}_v = 1 \tag{15}$$

Flow conservation constraints to find a path in the time-space graph:

$$\sum_{\ell \in \omega^+(v)} x_\ell - \sum_{\ell \in \omega^-(v)} x_\ell = -\mathrm{DST}_v \qquad v \in \delta^+(L^T) \tag{16}$$

$$\sum_{\ell \in \omega^+(v)} x_\ell - \sum_{\ell \in \omega^-(v)} x_\ell = 0 \qquad v \in V \setminus \delta^+(L^T) \cup \{v_\mathrm{O}\} \tag{17}$$

$$\sum_{\ell \in \omega^+(v_\mathrm{O}) \cup L^T} x_\ell = 1 \tag{18}$$

$$\sum_{\ell \in \omega^-(v_\mathrm{O})} x_\ell = 0 \tag{19}$$

Ensure the path of consist travel plan is not empty:

$$\sum_{\ell \in L^T} x_\ell \geq 1 \tag{20}$$

Ensure that the consist travel plan contains no critical locomotives after due date, i.e., maint_time, which was set to 90 calendar days in the numerical experiments (TimeDst(t) is the days from the start time of LAP plan period, to the arrival time of train t):

$$n_k \leq M \cdot (1 - x_\ell) \qquad \ell \in L^T, k \in K_c : \mathrm{TimeDst}(t) + m_{k_c} \geq \mathsf{maint\_time} \tag{21}$$

Ensure that each train in the consist travel plan has enough power:

$$\sum_{k \in K_r} \mathrm{HP}_k \cdot n_k + \sum_{k \in K_c} \mathrm{HP}_k \cdot n_k \geq \max_{\ell(\equiv t) \in L^T} (x_\ell \cdot \mathrm{HP}_t) \tag{22}$$

Ensure the lower & upper bounds of consist size:

$$\mathrm{consist\_size}_{\min} \leq \sum_{k \in K_r} n_k + \sum_{k \in K_c} n_k \leq \mathrm{consist\_size}_{\max} \tag{23}$$

The next set of constraints correspond to the linearization of objective (14).

$$y_{k,v} \leq n_k \qquad v \in V, k \in K \tag{24}$$

$$y_{k,v} \leq M \cdot \textsc{dst}_v \qquad v \in V, k \in K \tag{25}$$

$$y_{k,v} \geq n_k + M \cdot (\textsc{dst}_v - 1) \qquad v \in V, k \in K \tag{26}$$

Note that constraints (25) can be re-enforced as follows:

$$\sum_{k \in K} y_{k,v} \leq M \cdot \textsc{dst}_v \qquad v \in V \tag{27}$$

hence reducing the number of constraints as well.
The next set of constraints prevent two trains to be connected directly:

$$\sum_{\ell \in \omega^+(v_\text{O}) \cup L^T} x_\ell + \sum_{\ell \in \omega^-(v_\text{O}) \cup L^T} x_\ell \leq 1 \qquad v \in V \tag{28}$$

The master and pricing problems both can be solved by a general-purpose solver, e.g., Ilog Cplex.

2.5.4 Modified Model for Deadheading Analysis

The modification of LAP model in this section focuses on to reducing the deadheadings. First we change the MP's objective to reduce the total deadheading powers:

$$\min \quad \sum_{s \in S} \sum_{t \in s} z_s \cdot (\textsc{hp}_s - \textsc{hp}_t). \tag{29}$$

where $(\textsc{hp}_s - \textsc{hp}_t)$ represents that in each consist travel plan s, the sum of the power difference of assigned power and actually needed power for every train.

Then, add a new constraint in MP that limit total number of additional locomotives.

$$\sum_{v \in V^{\textsc{src}}} \sum_{k \in K} x_{kv}^{\textsc{need}} \leq \text{UB}^{\text{need}} \tag{30}$$

According to the MP's modification, the objective of PP has to be changed as:

$$
\begin{aligned}
\overline{\text{COST}} = \sum_{\ell^t \in L^T} \left(\sum_{k \in K} n_k \cdot \text{HP}_k - \text{HP}_t \right) \cdot x_{\ell^t} \quad &+ \quad \left[\sum_{k \in K_r} u^{(8)}_{kv} n_k + \sum_{k \in K_c} u^{(9)}_{kv} n_k \right] \\
- \sum_{k \in K_r \cup k_c} & \left(u^{(2)}_{kv_{\text{O}}} n_k - \sum_{v \in V \setminus (V^{\text{SRC}} \cup \{v^{\text{SINK}}\} \cup \delta^+(L^{\text{SHOP}}))} u^{(2)}_{kv} \cdot y_{k,v} \right) \\
& - \sum_{k \in K_r} \left(u^{(3)}_{k,v_{\text{O}}} \cdot n_k - \sum_{v \in \delta^+(L^{\text{SHOP}})} u^{(3)}_{k,v} \cdot y_{k,v} \right) \\
& - \sum_{k \in K_c} \left(u^{(4)}_{k,v_{\text{O}}} \cdot n_k - \sum_{v \in \delta^+(L^{\text{SHOP}})} u^{(4)}_{k,v} \cdot y_{k,v} \right) \\
& + \sum_{k \in K_r \cup k_c} \sum_{v \in V \setminus (V^{\text{SRC}} \cup \{v^{\text{SINK}}\} \cup \delta^+(L^{\text{SHOP}}))} u^{(5)}_{kv} \cdot y_{k,v} \\
& + \sum_{v \in \delta^+(L^{\text{SHOP}})} u^{(6)}_{k,v} \cdot y_{k,v} + \sum_{v \in \delta^+(L^{\text{SHOP}})} u^{(7)}_{k,v} \cdot y_{k,v} - \sum_{\ell \in L^T} u^{(11)}_t \cdot x_\ell. \qquad (31)
\end{aligned}
$$

The new objective is non-linear, in order to linearize it, we first add two decision variables: $Dead_{\ell^t} = \sum_{k \in K} n_k \cdot \text{HP}_k - \text{HP}_t$ represent the deadheading power for each train link $\ell^t \in L^T$.

$\alpha_\ell = Dead_\ell \cdot x_\ell$ It is used to linearize the product $Dead_\ell \cdot x_\ell$ of decision variables.

In addition, we add constraints in pp: first modify PP's Constraints (22) as

$$
\sum_{k \in K_r} \text{HP}_k \cdot n_k + \sum_{k \in K_c} \text{HP}_k \cdot n_k = x_\ell \cdot \text{HP}_t + Dead_\ell \qquad \ell \in L^T \tag{32}
$$

Then, to linearize the objective,

$$
\alpha_\ell \leq Dead_\ell \qquad \ell \in L^T \tag{33}
$$

$$
\alpha_\ell \leq M \cdot x_\ell \qquad \ell \in L^T \tag{34}
$$

$$
\alpha_\ell \geq Dead_\ell + M \cdot (x_\ell - 1) \qquad \ell \in L^T. \tag{35}
$$

Finally, the objective of PP is modified as:

$$
\begin{aligned}
\overline{\text{COST}} = \sum_{\ell\in L^T} \alpha_\ell \quad + \quad & \left[\sum_{k\in K_r} u^{(8)}_{kv} n_k + \sum_{k\in K_c} u^{(9)}_{kv} n_k\right] \\
& - \sum_{k\in K_r\cup k_c} \left(u^{(2)}_{kv_{\text{O}}} n_k - \sum_{v\in V\setminus(V^{\text{SRC}}\cup\{v^{\text{SINK}}\}\cup\delta^+(L^{\text{SHOP}}))} u^{(2)}_{kv}\cdot y_{k,v}\right) \\
& - \sum_{k\in K_r} \left(u^{(3)}_{k,v_{\text{O}}}\cdot n_k - \sum_{v\in\delta^+(L^{\text{SHOP}})} u^{(3)}_{k,v}\cdot y_{k,v}\right) \\
& - \sum_{k\in K_c} \left(u^{(4)}_{k,v_{\text{O}}}\cdot n_k - \sum_{v\in\delta^+(L^{\text{SHOP}})} u^{(4)}_{k,v}\cdot y_{k,v}\right) \\
& + \sum_{k\in K_r\cup k_c} \sum_{v\in V\setminus(V^{\text{SRC}}\cup\{v^{\text{SINK}}\}\cup\delta^+(L^{\text{SHOP}}))} u^{(5)}_{kv}\cdot y_{k,v} \\
& + \sum_{v\in\delta^+(L^{\text{SHOP}})} u^{(6)}_{k,v}\cdot y_{k,v} + \sum_{v\in\delta^+(L^{\text{SHOP}})} u^{(7)}_{k,v}\cdot y_{k,v} - \sum_{\ell\in L^T} u^{(11)}_t\cdot x_\ell. \quad (36)
\end{aligned}
$$

2.6 Numerical Results

2.6.1 Data Instances

Numerical experiments are generated from given train scheduling data and railway infrastructure of CPR. The data include train departure times and stations, train arrival times and stations, and horse-power requirements. And we use CPR's entire railway network (from Vancouver to Montreal, covering all of Canada and parts of the United States, shown in Figure 5), the number of each type of locomotives in operation, and the location and capacity of maintenance shops.

We use a set of 9 different types of locomotives, limiting our experiments to the most used locomotives in the CPR fleet of locomotives, as described in Table 1. As requested by the mathematical model, the number of types was doubled in order to distinguish the critical (about 20% of the overall number of locomotives) from the non critical locomotives.

We defined 11 test scenarios of increasing scheduled train size, from 113-train to 1,750-train scenarios. The largest scenario corresponds the typical number of CPR scheduled trains over a time period up to 2 weeks. The test scenarios are solved by CPlex 12.6.1 in a server with 40-cores, 1TB memory.

As the data of the locomotive location was not available, we use a heuristic in order to generate an initial geographical distribution of the locomotive fleet in the railway network.

Figure 5: CPR Railway Network[3]

2.6.2 Accuracy & Efficiency of the Solutions

In this section we discuss the quality of the LAP solutions. The results of each test scenario are listed in Table 2, from 113-train to 1,750-train scenarios. The first two columns provide the trains and time length of the test scenario. The third column shows the number of consist travel plans (columns) generated by LAP model, excluding those in the original input. The fourth column provides the number of consist travel plans selected in the final solution for the final ILP model. The fifth column shows the total number of locomotives demanded in operation. The sixth column shows the total additional locomotives needed

Model	Horsepower	units
GP38	3,000	191
GP40	3,000	83
GP40-2	3,000	92
SD40-2	3,000	323
SD60	3,800	251
SD90/4300	4,300	95
ES44AC	4,360	98
AC4400CW	4,400	1,160
AC4400CW-L	4,390	82
TOTAL	N/A	2,375

Table 1: Locomotives Types & Quantity

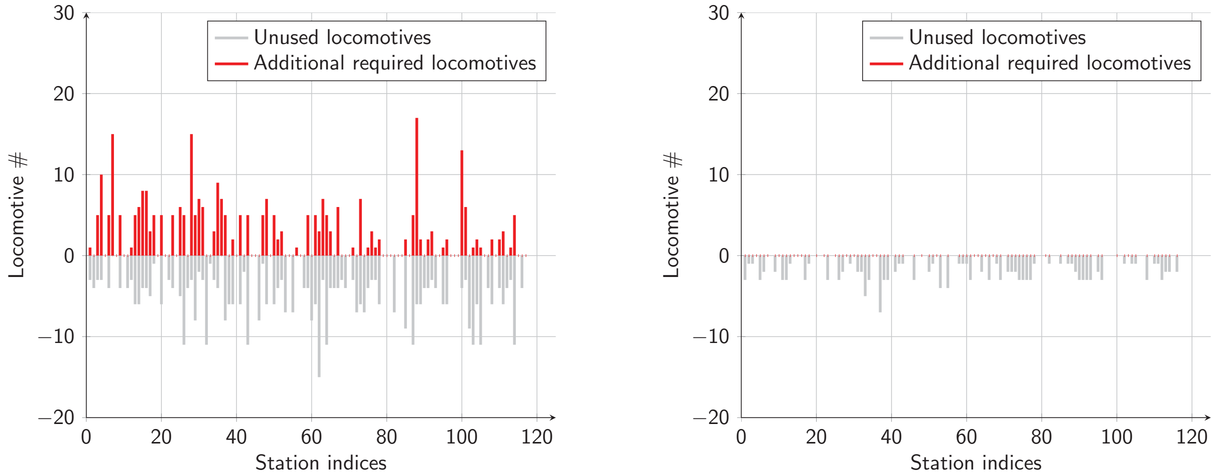
# Trains	Time Length	# Columns		# Locomotives		GAP (%)	CPU Time (hh:mm:ss)
		Generated	Selected	in Operation	Additional		
113	1 day	61	110	250	49	0.45	00:01:20
232	2 days	125	183	470	78	0.93	00:06:22
357	3 days	153	315	619	108	1.07	00:26:47
466	4 days	202	378	735	122	1.34	02:41:35
593	5 days	1,095	445	820	137	1.89	05:42:35
733	6 days	1,237	479	873	131	1.73	12:55:17
862	7 days	2,207	516	962	150	1.24	20:19:36
995	8 days	2,313	732	1,026	173	1.73	25:38:26
1,230	10 days	2,087	983	1,193	217	2.02	31:32:07
1,486	12 days	1,992	1,098	1,225	258	1.80	37:12:25
1,750	14 days	1,835	1,294	1,289	306	1.82	44:41:53

Table 2: Results in Different Number of Trains

for the solution. The seventh column shows the gap between the results from ILP and LP formats. We can see that the gap for each scenario is no more than 2.1%. The last column gives the computational times. We can observe that we can solve locomotive assignment problems with up to about 1,750 trains with all the real-world constraints, e.g. shop and consist busting in a acceptable time.

We expect that, with the addition of a better heuristic in order to generate an initial set of consists and locomotive assignments, we can significantly reduce those computing times, and therefore solve larger instances. This would fully answer their need of an automated tool for optimizing their locomotive assignment, and tentatively determining the optimal size of their locomotive fleet.

From the results in Table 2, especially in sixth column for total additional locomotives needed, we can see that the initial distribution of the locomotives has some issues. For example, as Figure 6(a) shows for 1,750-train scenario, which is exacted from the train schedule given by CPR, the LAP result needs some additional locomotives in some stations, and also leaves some locomotives out of operation. Based on the analysis of solution, we re-initialize the input, adjust the distribution of the locomotives, and use it as the input of LAP model. The new result for 1,750-train scenario is shown in Figure 6(b). With the distribution-modified initial input, we eliminate the additional locomotives, reduce the number out of operation, without lower the solution quality. So the LAP model can work also for the locomotive distribution area.



(a) Before Optimization

(b) After Optimization

Figure 6: Locomotive Initial Position for 1,750 Train Scenario

2.6.3 Analysis of In-Consist Waiting Time

In this section we analyze the in-consist waiting time. Figure 7 shows the train configuration in each consist travel plans, in the result for 1,750-train scenario. The horizonal axis represents the serial number of consist travel plans (the one-consist travel plans are dismissed), the vertical axis represents the time in days of the whole LAP scheduling period. Each vertical bar represents a consist travel plan from the departure time of its first train to the arrival time of its last train. The read parts of the bar represent the trains of consist travel plan, and the grey intervals are the in-consist waiting times.

We can see that, in most of the consist travel plans, the consist running time is much more than the in-consist waiting time, that reflects the efficiency of the solution. The few exceptions, e.g., consist travel plans #50, contains at least one local train, i.e., the short distance trains with less than 3 hours running time.

2.6.4 Analysis of the Power Assignment

Here we discuss the unused power and/or deadheading locomotives in each consist travel plan. Figure 8 shows power assigned of each train in consist travel plan (the one-consist travel plans are also dismissed), in the result for 1,750-train scenario. First, in Figure 8(a), we analyze the requested and assigned power. The horizonal axis represents the serial number of consist travel plans, the vertical axis represents the number of trains in the consist travel plan. Each vertical bar represents a consist travel plan, and each unit is a train. In each unit, the red part represents the ratio of requested power of each train in the consist travel

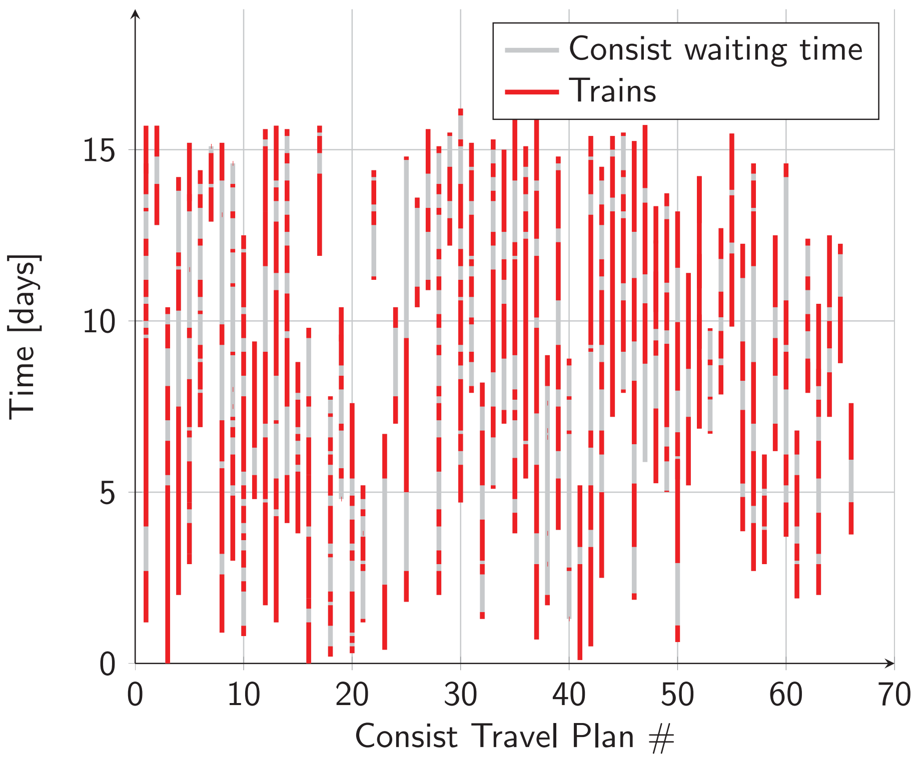


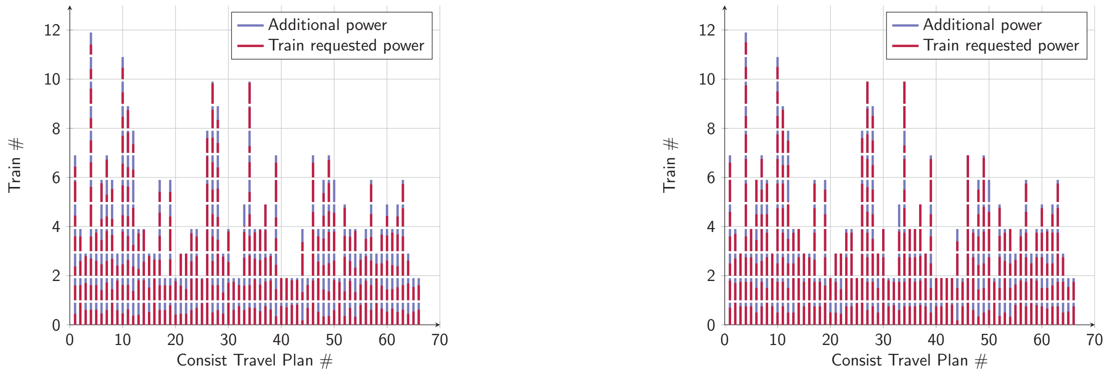
Figure 7: Consist Waiting Time in 1,750 Train Scenario

plan to the assigned power which is normalized to 1 unit, and the blue part is the unused power.

We know that the unused power can be the contribution of deadheading locomotive, but not all for them. For example, train A requires 10,000HP, and is assigned with a consist of 4 locomotives of 4,400HP each. So, we can see that 1 deadheading locomotive is deadheading, but there is still another 3,200HP assigned but unused left. So, in Figure 8(b), we analyze the number of deadheading locomotives vs. total consist. The difference from Figure 8(a) is that the red part represents the ratio of locomotives in operation of each train in the consist travel plan to the total assigned consist which is normalized to 1 unit, and the blue part represents the deadheading locomotives. In this figure we can see that most trains do not have deadheading locomotives for the result.

2.6.5 Results for Modified Model for Deadheading Analysis

The result is shown as the curves in Figure 9. The read curve represents the number of deadheading locomotives vs. the size of relaxation (additional locomotives). The blue curve represents the number of consist travel plans vs. the size of relaxation We can see that as we relax the total number of locomotives in operation, the deadheading locomotives



(a) Power Assigned of Each Train in Consist Travel Plan

(b) Deadheading Locomotives of Each Train in Consist Travel Plan

Figure 8: Results for 1,750 Train Scenario: Train Strings with Multiple Trains

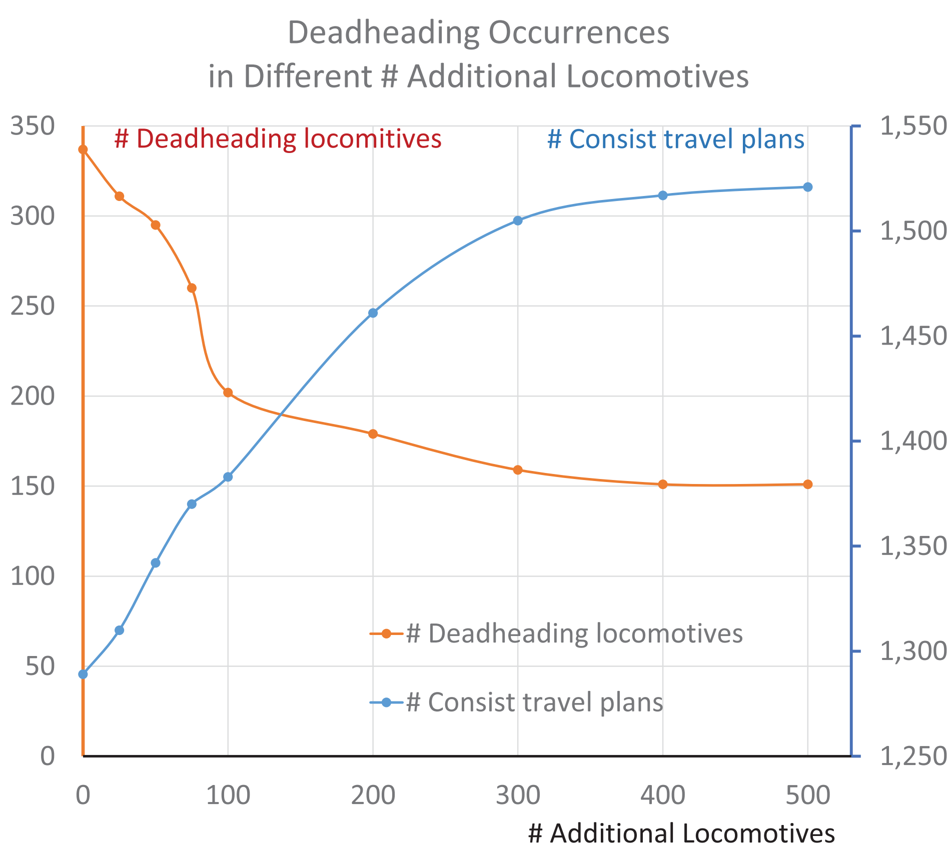


Figure 9: Deadheading Occurrence in Different Number of Additional Locomotives

reduces, but after the relaxation (additional locomotives) reach 300, we can not decrease the deadheading much. In the same time, the number of consist travel plans raises concurrently, until the relaxation reach 300. So we have to make a compromise between minimization of locomotives, and reducing the waste of deadheading.

2.7 Conclusion

We proposed a new mathematical model for optimizing locomotive assignment to a set of scheduled trains. It allows the *exact* solution of larger data instances than reported so far in the literature in a reasonable amount of time, as a planning tool. In addition, we believe that the speed of the solution process can be easily reduced with the addition of a more powerful heuristic in order to generate an initial solution, while preserving the generation of an exact locomotive assignment.

Chapter 3

An Enhanced Decomposition Scheme for Locomotive Assignment Problem

H. Tian and B. Jaumard. submitted for publication, 2017. An extended abstract of this paper has been published in 16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016). [33]

3.1 Introduction

Comparing to other transportation methods, e.g., via road truck or cargo airplane, railway transportation has its advantages of low energy consumption and high safety, so to attract concentration in academic and industry areas. In those areas locomotive management is one of the major problems, for that the high capital investment on the equipment purchase and the operation expense.

The management of locomotive assigns proper locomotives to trains in schedule, in order to satisfy power and other requirements. The **consist** is created by a group of locomotives to tract the train. To match the different request of other train(s), the consist may be broken to re-assign its locomotives to other trains. This time-consuming process is called **consist busting**. The railway industry wants to avoid this procedure as much as possible for the additional labor & operational cost and time, and also for the loss of robustness of the train/locomotive schedule. Because consist busting allows an outbound train to get locomotives from multiple inbound trains, if any of the source inbound trains is delayed, the outbound train has to be delayed as well.

For locomotive relocation, locomotive(s) can be attached to the consist of a train, and be driven to another station. This process is called **deadheading**. A deadheading locomotive operates under its own power but does not pull the train. The deadheading operation is profitless but is an important method for the flexible assignment of locomotive. Another way of relocating is light traveling, which allows grouping of several locomotives with only one of them for power pulling. Usually the railway company does not use light travel for the extra affect on the train schedule. In this paper we only consider deadheading. Another time consuming but compulsory process is the locomotive maintenance, which require each locomotive to be maintained in a shop based on a regular time schedule (usually every 90 days). A locomotive due to maintenance is called **critical** and should not be used to pull a train after the due date.

Our work concentrates on the optimization of locomotive assignment problem (LAP), which is to optimize the total expense of locomotive fleet for a given train schedule under the constraints of the horsepower requests and the other technical and business constraints.

This paper is designed as follows. In Section 3.2, we take a review of the literatures for LAP modeling and for the acceleration of column generation. Section 3.3 generally describe the LAP. Section 3.4 gives the details of LAP model. In Section 3.5, the enhanced solution process (**multi-CG**) for the model is presented. And in Section 3.6, we provides the numerical results and analysis.

3.2 Literature Review

3.2.1 Previous LAP Models

There exist some research works for the optimization of locomotive assignment problem, which can be classified to exact mathematical models and heuristics. Our work focuses on the exact optimization model so here we lists the review of the works on this area. The works in heuristics has been reviewed in the survey of Piu *et al.* [45].

Ziarati *et al.* [55] reformulated LAP as an integer multi-commodity network flow problem which has a classical model of mixed integer linear problem (MILP) with the nonlinear part for maintenance constraints. Then they decomposed the model into a column generation (CG) format, which is solved by a branch and price algorithm. To avoid the scalability issue in computational time, Ziarati *et al.* merged several sub-problems of 2 or 3 days into a feasible solution for full model of 1-week long. Their model can handle 1,249 locomotives of 26 types including 171 critical ones, and satisfies 1,988 train-segment requests, over the railway network of Canadian National North America which contains 26 stations, 164 outposts and 18 shops. The authors claim that their decomposition methods have a better locomotive schedule than the CN's. This model does not consider consist busting issue.

Rouillon *et al.* [48] improveed the solution algorithm of Ziarati *et al.* [55] with different branching methods and search strategies to develop a branch-and-price algorithm for LAP of a freight railway on operational level.

Ahuja *et al.* [9] developed a MILP model of LAP for CSX Transportation. They also formulate LAP as a locomotive flow model, i.e., an integer multi-commodity flow problem. The network is similar to the one of Ziarati *et al.* [55], but the train schedule and the locomotive assignment plan is assumed to be cyclic every week. However, the maintenance process, i.e., forcing the locomotive regularly back to shop site is not considered, neither for the consist busting issue. The model has been validated with CSX data which contains 3,324 trains, 119 stations and 3,316 locomotives of 5 types. There is no CG reformulation of the MILP model, instead, the authors develop a neighborhood search heuristic to improve the performance for large scale data instances, with no information on the accuracy of the output solutions.

The models of Ziarati *et al.* [55] and Ahuja *et al.* [10] have several issues. First, the high rate of consist busting (the medium rate is 50%) in the solution which means it tends not to assign the same consist to the same train on different days. The second is that their MIP models have severe scalability issues when it comes to solve real-life data instances. Even for generating a first feasible solution, the convergence is slow: for some input data sets for the model of Ahuja *et al.* [10], Vaidyanathan *et al.* [54] said that in several hours there is still no convergence.

To address the issues of the model of Ahuja *et al.* [10], Vaidyanathan *et al.* [54] extended Ahuja's model to a consist-based one. The configurations of consists are generated by a pre-processing algorithm. Then, instead of assigning locomotives, the model assigns consists to pull the scheduled trains with respect to the minimum power and other business constraints. Their consist-based formulation uses a data set with 382/388 trains, 6 locomotive types, 87 stations, and 3 up to 17 types of consists in the test scenarios.

The potential issue of the consist-based formulation is: the greater optimization of the solution requires greater numbers of configuration types of consist. However, the computational time will grow as well, and even faster (it is true for normal MILP, but can be solved by CG decomposition as we proposed in next section). The other issue is that Vaidyanathan *et al.* [54] do not consider the maintance/shopping constraints for locomotives, and only assume the locomotive plan is cyclic every week. In addition, their consist-based model does not take the initial locomotive positioning into account. Instead, the authors propose a post-processing algorithm which minimizes the locomotive repositioning (light trains) in order to ensure a smooth transition from the current locomotive locations to the locomotive location requirements in the solution of the LAP model. This entails a lot of locomotive reposition issues.

There are research works to solve the problem different but similar to LAP. Cordeau *et al.* [22] solved the locomotive and car assignment in passenger transportation, with an exact model based on the Benders decomposition approach. The model is tested over a local part of VIA Rail Canada network including 9 main stations, with 300 trains, 2 types of locomotives and 4 types of cars. But this mode does not consider maintenance constraint, neither for consist busting issue.

Fügenschuh *et al.* [26] focused on locomotive and car cycle scheduling problem with time window. Their integer linear programming (ILP) model does not consider the maintenance and consist busting, but allow the train delay within given time window. It has been tested by 120 trips, 43 locomotives of 4 types on a 1,919-km railway networks. This mode does not consider maintenance constraint, neither for consist busting issue.

Cacchiani *et al.* [17] proposed two ILP formulations on the train unit assignment problem in passenger transportation, in which a set of passenger cars with a supported locomotive is self-contained and respond for one or some trips, without the maintenance constraint, neither for consist busting consideration. One ILP model is solved by linear programming (LP) relaxation based heuristic, the other by Lagrangian relaxation. For the problem up to 660 trips, 75 train units of 10 types, the Lagrangian relaxation based heuristic works better than the LP relaxation.

Our previous work Jaumard *et al.* [35] proposed a consist travel plan (previously called train string) based optimization model with maintenance constraint and consideration of consist busting. With column generation decomposition, the model solved LAP with up to 1,394 scheduled trains and 9 types of locomotives in the entire Canada Pacific Railway network. But computational time for the largest data set takes more than 2 days.

3.2.2 Acceleration Strategies for CG and Similar Algorithms

From the previous section, some models including ours applied Column Generation (CG) decomposition technique to solve the scalability issue of high time and memory requirements. But the requirement of time is still enormous, e.g., several days. To solve the computational time and convergence rate issues, there are strategies in CG's different stages.

In problem pre-process stage, some heuristics are developed for a better worm start, e.g., to reduce the initial network size(e.g., Mingozzi *et al.* [42]), to generate a good-enough initial solution (e.g., Sadykov *et al.* [50]), to decompose a large problem into parts to lower the time and memory requests, and merge the sub-solutions after (e.g., Desaulniers *et al.* [24]).

In the sub-problem stage, Chen *et al.* [19] use some problem-specific knowledge to generate a column-pool a priori for the subproblem, and allow it to select solutions from the pool.

In column generation practice, some schemes allow a subproblem (usually a heuristic) to return multiple columns with negative reduced cost (e.g., Goffin *et al.* [30]).

Other possible strategies are available in relaxing time/cost interval, node sorting, partial pricing network, subproblem aggregating, etc..

In master problem stage, Surapholchai *et al.* [51] develop Eligen-algorithm that applies column elimination which removes columns with positive reduced cost from the matrix. Saddoune *et al.* [49] use dynamic constraint aggregation to reduce number of constraints and reintroduce them as needed are two general strategies. Sadykov *et al.* [50] use a diversified diving heuristic to get feasible and good integer solution.

3.3 Problem Statement

Locomotive assignment problem (LAP) deals with a locomotives fleet and the assignment of those locomotives to scheduled trains. The optimization of LAP is to minimize the total expense of the locomotives that satisfy the scheduled trains with given constraints, e.g., horsepower, minimum locomotive number, and maintenance requirements, etc.. Our LAP deals with 9 different type of locomotives, each type has two statuses or sub-types (regular/crytical) for the maintenance. The LAP model regards locomotives of the same sub-type as the items with no difference.

Maintenance Each locomotive needs to be maintained of a regular time period (here assumed to be a calendar interval, see, e.g., Railway Locomotive Inspection and Safety Rules [5] for more details on maintenance intervals.). A locomotive is called critical when it is due to maintenance. A critical locomotive must stop at a shop for maintenance operations. After the process it will be regular one.

Consist Travel Plan A consist travel plan is defined as a set of trains that use the same locomotive consist one train after the other one, without any consist busting. A train in schedule should be covered by one and only one plan. Consist travel plans (and shop links) must be spaced a minimum time (2 hours in our numerical experiments) in order to allow consists to be busted and reassembled. Within a consist travel plan, the time difference of two consecutive train is spaced by a minimum time period, at least 1 hour in our numerical experiments.

Multi Commodity Network Similar to some LAP models, we regards LAP as a multi-commodity network problem, see Figure 10. The node v is associated with two parameters: LOCATION(v), corresponding to a railway station location, and, TIME(v), which is the time of the beginning or the end of an activity, and is expressed in minutes. The arcs represent locomotive process such as pulling train from station to station, waiting for next assignment, or maintenance.

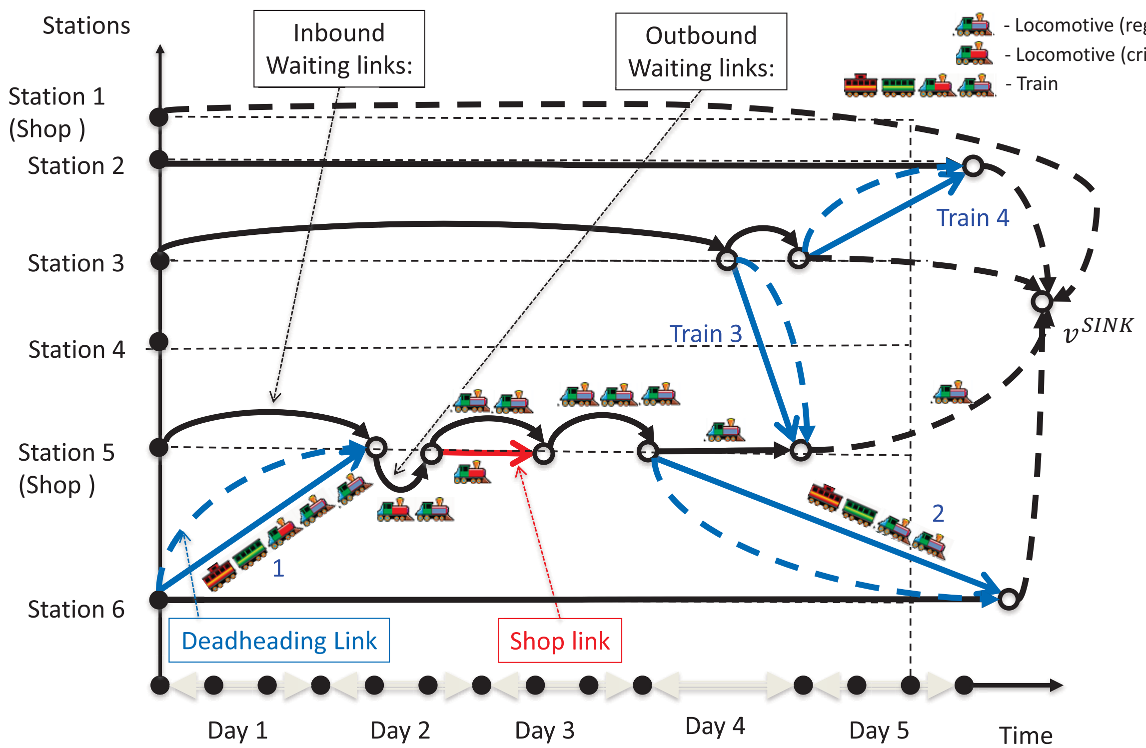


Figure 10: Multi commodity network

In the generic multi-commodity network $G = (V, L)$, V represents the set of nodes, indexed by v:

V^{SRC}: the set of sources nodes v^{SRC} where at the start of the planning period, some locomotives are ready for assignment.

v^{SINK}: an artificial node for the end of the network flows.

L is the set of arcs ℓ. $L = L^T \cup L^{\text{SHOP}} \cup L^W \cup L^D$, which represent train links, shop links, waiting links and deadheading links respectively.

Waiting Links A waiting link connects two train links together, and is defined as a link with its two endpoints associated with the same location (station) components. To ensure the time constraints between, e.g., two successive train departures, or for reassembling a consist, we divide the waiting links into the inbound and the outbound waiting links: $L^{\text{W}} =$

$L^{\text{W\_IN}} \cup L^{\text{W\_OUT}}$ with $L^{\text{W\_IN}} \cap L^{\text{W\_OUT}} = \emptyset$. In addition, such a division will allow us to identify the consist busting. An ***inbound waiting link*** ($\ell^w \in L^{\text{W\_IN}}$) starts at the destination node of a train link or at a source node, and ends at the nearest origin node of another train/shop link, with a time duration that is at least the time required to re-assign a given consist to another train. An ***outbound waiting link*** ($\ell^w \in L^{\mathbf{w\_out}}$) starts and ends at the nearest origin nodes of two different train links, or at an origin node and the dummy sink node, without any time restriction. Figure 10 illustrates the division of waiting links. Shop links are considered as train links, with respect to the definition of inbound/outbound waiting links.

3.4 LAP Model

3.4.1 Notations

s: consist travel plans : a list of trains in LAP solution that can be pulled by the same locomotive consist.
k: a certain type of locomotive with several parameters, e.g., the horsepower HP_k. k_r, k_c: the subtype of k for regular/critical.
K: the set of types of locomotives.
n_k^s represents the number of locomotives in type k belongs to the consist travel plan $s \in S$.
$n_{k,v}^{\text{SPARE}}$ represents the number of spare locomotives in type k in start node $v \in V^{\text{SRC}}$.
$d_\ell^s = 1$ if train link $\ell \in L^T$ belongs to consist travel plan s, 0 otherwise. Note that d_ℓ^s is not a decision variable, but an attribute of consist travel plan s.
$\text{CAP}(\ell^{\text{SHOP}})$ = upper bound of critical locomotives that can be maintained in shop link $\ell^{\text{SHOP}} \in L^{\text{SHOP}}$.
$\text{TimeSrc}(t), \text{TimeDst}(t)$ = the start and end time (in days) of train t, counted from the start time of LAP scheduling period.

3.4.2 Variables.

We use three sets of variables:
$z_s = 1$ if consist travel plan s is selected, 0 otherwise.
x_{kv}^{NEED} = number of additional required locomotives of type k at source node $v \in V^{\text{SRC}}$ in order to be able to assign adequate locomotives to all trains.
$x_{k\ell}^{\text{LOCO}} = 1$ if locomotive k goes through link $\ell \in L^W \cup L^D \cup L^{\text{SHOP}}$, 0 otherwise.

3.4.3 Objective

The objective of LAP is to minimize the total number of locomotives required, with the lowest consist busting occurrences. Since the two components are poles apart, we make the compromise of the objective:

$$
\begin{aligned}
\min \quad & \sum_{\ell \in \omega^-(v^{\text{SINK}})} \sum_{k \in K} \text{PENAL}_k \cdot x_{k\ell}^{\text{LOCO}} \\
+ \quad & \sum_{\ell \in L^D} \sum_{k \in K} \text{PENAL}_k \cdot x_{k\ell}^{\text{LOCO}} \\
+ \quad & \sum_{v \in V^{\text{SRC}}} \sum_{k \in K} \text{PENAL}_k \, x_{kv}^{\text{NEED}} \\
+ \quad & \sum_{s \in S} \sum_{k \in K} n_k^s \, z_s
\end{aligned} \tag{37}
$$

3.4.4 Constraints

$$
\begin{aligned}
& \sum_{s \in S_v^+} n_k^s \, z_s + \sum_{\ell^w \in \omega^+(v)} x_{kL^{\text{W}}}^{\text{LOCO}} + \sum_{\ell^w \in \omega^+(v)} x_{k\ell^D}^{\text{LOCO}} - x_{kv}^{\text{NEED}} \\
& \leq n_{k,v}^{\text{SPARE}} \qquad\qquad k \in K_r, v \in V^{\text{SRC}}
\end{aligned} \tag{38}
$$

$$
\begin{aligned}
& \sum_{s \in S_v^+} n_k^s \, z_s + \sum_{\ell^w \in \omega^+(v)} x_{kL^{\text{W}}}^{\text{LOCO}} + \sum_{\ell^w \in \omega^+(v)} x_{k\ell^D}^{\text{LOCO}} \\
& \leq n_{k,v}^{\text{SPARE}} \qquad\qquad k \in K_c, v \in V^{\text{SRC}}
\end{aligned} \tag{39}
$$

$$
\sum_{\ell \in \omega^-(v^{\text{SINK}})} x_{k\ell}^{\text{LOCO}} \leq n_k \qquad\qquad k \in K \tag{40}
$$

$$\sum_{s\in S_v^+} n_k^s\, z_s + \sum_{\ell\in\omega^+(v)\,\cap\,(L^{\text{WAIT}}\cup L^D)} x_{k\ell}^{\text{LOCO}}$$
$$= \sum_{s\in S_v^-} n_k^s\, z_s + \sum_{\ell^w\in\omega^-(v)\,\cap\,(L^{\text{WAIT}}\cup L^D)} x_{k\ell}^{\text{LOCO}}$$
$$v \in V \setminus \left(V^{\text{SRC}} \cup v^{\text{SINK}} \cup \delta^+(L^{\text{SHOP}})\right),$$
$$k \in K_r \cup K_c \tag{41}$$

$$\sum_{s\in S_v^+} n_{k_r}^s\, z_s + \sum_{\ell\in\omega^+(v)\,\cap\,(L^{\text{WAIT}}\cup L^D)} x_{k_r\ell}^{\text{LOCO}}$$
$$= \sum_{s\in S_v^-} n_{k_r}^s\, z_s + \sum_{\ell\in\omega^-(v)\,\cap\,L^{\text{SHOP}}} x_{k_c\ell}^{\text{LOCO}}$$
$$+ \sum_{\ell\in\omega^-(v)\,\cap\,(L^{\text{WAIT}}\cup L^D)} x_{k_r\ell}^{\text{LOCO}}$$
$$v \in \delta^+(L^{\text{SHOP}}), k = \{k_r, k_c\} \in K \tag{42}$$

$$\sum_{s\in S_v^+} n_k^s z_s + \sum_{\ell\in\omega^+(v)\,\cap\,(L^{\text{WAIT}}\cup L^D)} x_{k\ell}^{\text{LOCO}}$$
$$= \sum_{s\in S_v^-} n_k^s z_s + \sum_{\ell\in\omega^-(v)\,\cap\,(L^{\text{WAIT}}\cup L^D)} x_{k\ell}^{\text{LOCO}}$$
$$v \in \delta^+(L^{\text{SHOP}}), k \in K_c \tag{43}$$

$$\sum_{s\in S_v^+} n_k^s\, z_s \le \sum_{\ell^w\in\omega^-(v)\,\cap\,L^{\text{WAIT}}} x_{k\ell}^{\text{LOCO}}$$
$$v \in V \setminus \left(V^{\text{SRC}} \cup v^{\text{SINK}} \cup \delta^+(L^{\text{SHOP}})\right),$$
$$k \in K_r \cup K_c \tag{44}$$

$$\sum_{s\in S_v^+} n_{k_r}^s\, z_s \le$$
$$\sum_{\ell\in\omega^-(v)\,\cap\,L^{\text{SHOP}}} x_{k_c\ell}^{\text{LOCO}} + \sum_{\ell\in\omega^-(v)\,\cap\,L^{\text{WAIT}}} x_{k_r\ell}^{\text{LOCO}}$$
$$v \in \delta^+(L^{\text{SHOP}}), k \in K \tag{45}$$

$$\sum_{s \in S_v^+} n_k^s z_s \leq \sum_{\ell \in \omega^-(v) \cap L^{\text{WAIT}}} x_{k\ell}^{\text{LOCO}}$$
$$v \in \delta^+(L^{\text{SHOP}}), k \in K_c \quad (46)$$
$$\sum_{s \in S} d_\ell^s \cdot z_s = 1 \qquad \ell \in L^T \quad (47)$$
$$\sum_{k \in K} x_{k_c \ell^{\text{SHOP}}}^{\text{LOCO}} \leq \text{CAP}(\ell^{\text{SHOP}}) \qquad \ell^{\text{SHOP}} \in L^{\text{SHOP}} \quad (48)$$
$$\sum_{k \in K} x_{kL^{\text{W}}}^{\text{LOCO}} = 0$$
$$\ell^w \in L^{\text{W\_IN}} \setminus \omega^+(V^{\text{SRC}}) : \mathsf{time}(\ell^w) < \mathsf{dwell\_loco}. \quad (49)$$

Constraints (38), (39), and (40) allows the model to use additional locomotives x_{kv}^{NEED} for each source node to guarantee the feasible solution. Constraints (41), (42) and(43) are the flow conservation constraints except for the source and dummy sinking nodes. Note that critical locomotives are relabelled to regular after the maintanceshop process. This relabelling will be taken care thanks to special flow conservation constraints at the shop end nodes in the proposed LAP model. In addition, the flow conservation constraints above allows that two consist travel plans are connected directly to reuse (part of) the same locomotives. Constraints (44), (45), (46) guarantee to avoid this issue. Constraints (47) limit each train to be covered by one and only one consist travel plan. Constraints (48) guarantee the limit of shop capacity. Constraints (49) guarantee the minimum time for locomotives process between two consecutive consist travel plans.

3.5 Solution Process

3.5.1 CG Decomposition

The model we proposed has a warm start with a limited number of consist travel plans as a initial solution. The additional improving consist travel plans are generated by a consist travel plan generator, so-called Pricing Problem (PP) (Chvátal *et al.* [20]) corresponding to the Restricted Master Problem (RMP) representing the original model. Each iteration RMP gets its LP solution, the dual values and a source node are passed to PP, which generate an optimal consist travel plan (column) with negative reduced cost and add it back to RMP for next iteration. The CG process continues until PP can not find any column with negative cost, and then RMP will get the ϵ-optimal solution of ILP. The details of this column-generation (CG) decomposition process are available in our previous paper [34].

In the CG decomposition process, we get the optimal solution but the computational time is still high with the largest data sets. By the study of the pricing problem, we find out two potential areas for improvements.

First one is that for each pricing problem, even with a given origin node, it need to check the whole time-space network. Second and the most important area is, to stop the CG process with the optimal solution after the last solution with negative reduced cost, we need to check the whole origin node set to confirm that each choice has no new column generated.

For the two potential improvements we proposed two enhanced processes for LAP model in next two sections.

3.5.2 Enhanced Pricing Problem: Preprocessing for Conflict Graph

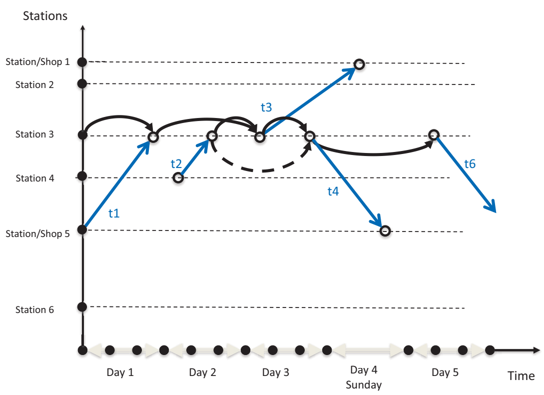
For the first area described above, we provide two ways for the improvement of reductions: the first one is to cut off the redundant links from the time-space network architecture for the pricing problem, and the second one is to add cuts into the pricing problem that un-connectable trains are mutual exclusive. Both ways need for a key point: the conflict graph for trains.

Similar to Barber *et al.* [13], the conflict graph of LAP is the set of trains that can be connected by the waiting links, i.e., those trains can be assigned to a consist travel plan or consist travel plan.

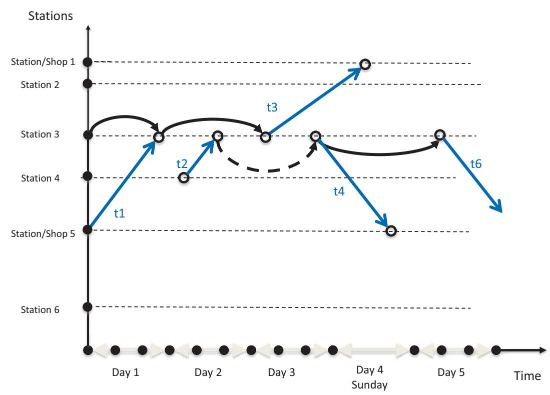
Reduce Network Architecture The pricing problem of LAP needs the input of the whole time-space network and the initial starting node which actually is the source of a selected train (link). By the given train link, we can find out that not the whole network transferred from RMP is needed by PP, we can remove part of the links (train links and waiting links) to reduce the network size so that reduce the size of PP. It is a three-step process.

Step 1: Restricted Network Architecture for Master and Pricing Problems In our previous paper, and also in Section 3.3, one of the two types of waiting links, the so-called An ***inbound waiting link***, has two lower bounds for the time duration in the master and pricing problem: on one hand, in RMP, it should be no less than 2 hours, with respect to the time consumption for consisting busting process; the other hand, in PP, the lower bound is only 1 hour, which is the lowest time duration for the inbound waiting link.

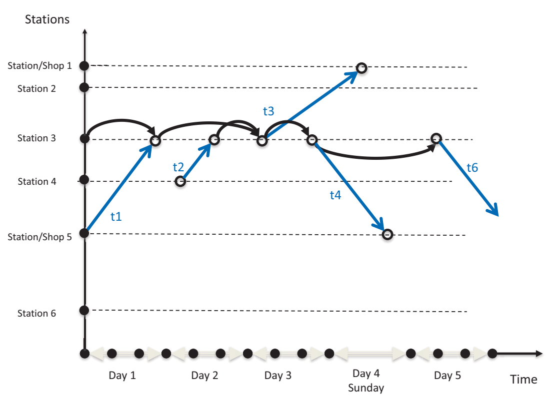
During the processing for LAP, which the input is only the given scheduled trains, we generate the waiting links for each given train $t \in T$:



(a) Original Time-Space Network



(b) Waiting Links for RMP



(c) Waiting Links for PP

Figure 11: Preprocessing: Reduce Network Architecture

1. from train t, find the next nearest train t' in the same station with at least 1 hours time later, generate a waiting link ℓ' from t to t', e.g., in Figure 11(a), the waiting link connects trains t_1 to t_3, and t_2 to t_3.

2. if t' is at least 2 hours later than t, e.g., waiting link connects trains t_1 to t_3, stop.

3. otherwise find the next nearest train t" in the same station with at least 1 hours time later, generate another waiting link ℓ" from t to t", e.g., the new waiting link connects trains t_2 to t_4.

4. connect the nearest pair of source nodes of the two outbound trains, e.g., waiting link connects trains t_3 to t_4.

We can see that, if we generate both ℓ' and ℓ", e.g., from trains t_2 to t_3, t_2 to t_4, and from t_1 to t_3, there exist transitive links which means redundancy. And since in RMP, we only use the inbound waiting links no less than 2 hours, in PP we use the time limit of 1 hour, we can remove the transitive links for different environment to reduce the size of the network. So in Figure 11(b), we get the reduced network for RMP, and in Figure 11(c), we get the reduced network for PP.

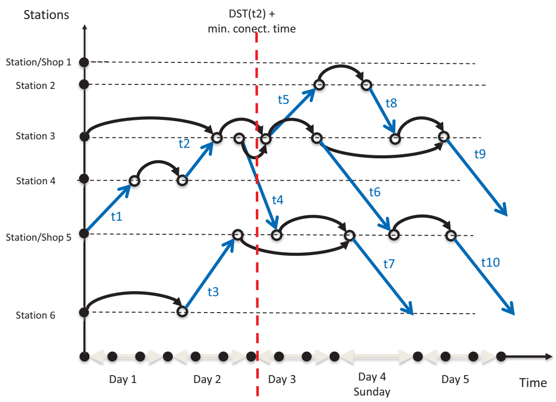
Step 2: Time Confliction The first step is easy, that the train (links) that has time conflict with the initial train should be removed. As the Figure 12 shows: in Figure 12(a), there is the original time-space network, the input train is t_2. So the time bound can be set to the arrival time of t_2 plus the inimical connection time for consist process, as the red vertical dash line shows. In Figure 12(b), the trains that conflict with train t_2 are detected: t_1, t_3, t_4.

Step 3: Connection Confliction The second step is that based on the network from step 1, check the connection of any train links, so we can remove the trains that are un-reachable from the given initial train. We develop an algorithm to generate the conflict graph, as shows in the Algorithm 1.

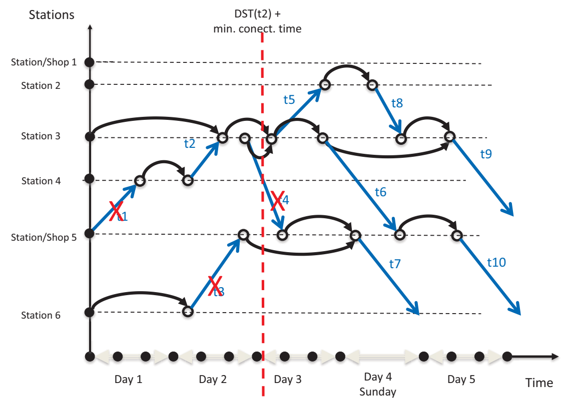
Based on the algorithm, we detect that in the network in Figure 12(c), train t_7 is un-reachable so can be removed. The final reduced time-space network (conflict graph) for PP which origin is train t_2 is showed in Figure 12(d). We also noticed that some conflict graphs are subset of other one. Those conflict graphs are redundant and will be removed in the pre-process.

3.5.3 Enhanced PP: Multiple Column Generation

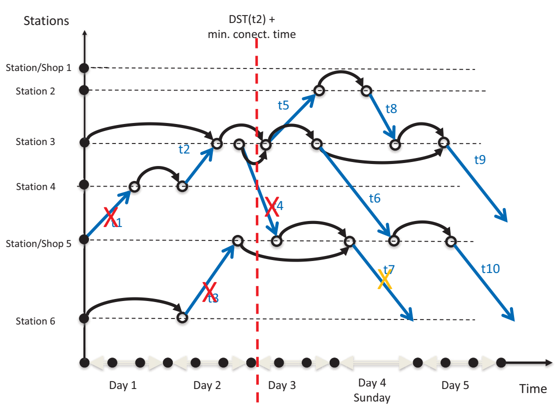
In normal CG process, the pricing problem (PP) uses the same data set of RMP and this dual value set to get the optimal solution. In our LAP model, it means PP has the flexibility



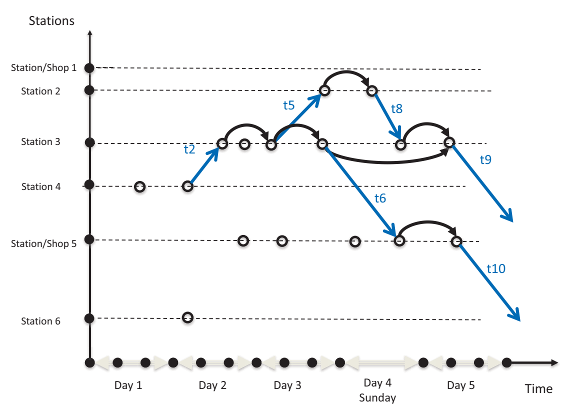
(a) Original Time-Space Network



(b) Remove Time Conflict Trains



(c) Remove Connection Conflict Trains



(d) Final Reduced Time-Space Network

Figure 12: Preprocessing: Reduce Network Architecture

Algorithm 1 Algorithm for Preprocessing: Conflict Graph

Require: time-space network $G = (V, L)$, $L = L^T \cup L^W$
Require: $MAP\{\ell : L_\ell^{\text{rightHandConnected}}\}$
$MAP\{\ell : L_\ell^{\text{trainConnectable}}\} \leftarrow \varnothing$
$MAP\{\ell \in L^T : L_\ell^{\text{trainConnectable}}\} \leftarrow \varnothing$
function GETCONNECTABLETRAINS(ℓ)
  if $\ell \in KEY(MAP\{\ell : L_\ell^{\text{trainConnectable}}\})$ **then**
    do nothing
  else if SRC(ℓ) == DummySinkNode **then**
    $MAP\{\ell : L_\ell^{\text{trainConnectable}}\} \leftarrow [\ell : \varnothing]$
  else
    for all $\ell' \in Value\left(MAP\{\ell : L_\ell^{\text{rightHandConnected}}\}, \ell\right)$ **do**
      GETCONNECTABLETRAINS(ℓ')
      $L_\ell^{\text{trainConnectable}} \leftarrow$
        $L_\ell^{\text{trainConnectable}} \cup \{\ell'\} \cup Value\left(MAP\{\ell : L_\ell^{\text{trainConnectable}}\}, \ell'\right)$
    end for
    $MAP\{\ell : L_\ell^{\text{trainConnectable}}\} \leftarrow \left[\ell : L_\ell^{\text{trainConnectable}}\right]$
  end if
end function
for all $\ell \in L$ **do**
  GETCONNECTABLETRAINS(ℓ)
  if $\ell \in L^T$ **then**
    $MAP\{\ell \in L^T : L_\ell^{\text{trainConnectable}}\} \leftarrow$
      $\left[\ell : Value\left(MAP\{\ell : L_\ell^{\text{trainConnectable}}\}, \ell\right)\right]$
  end if
end for
return $MAP\{\ell \in L^T : L_\ell^{\text{trainConnectable}}\}$

to choose any train source node as the origin to build the consist travel plan as the column for RMP. But, it take a very long time for the solution for only one column. So in our original LAP with CG and the modified version in Section 3.5.2 we choose to fix the origin node of PP to limit its flexibility so as to speed up the solution time. The following issue is we have to check the whole round robin order of the origin node set to stop CG process properly.

Based on the idea of conflict graph we discussed, we have another choice: instead of fixing the origin node in round robin, we allow PP to select one reduced sub-network, and create an optimal column from it. So if PP can not generate a new column with negative reduced cost, the CG process stops.

In addition, we allow PP to consider the whole network with conflict graphs, and generate a column with reduced cost from each graph. In those generated columns, one is the best of the total network, and each of the others is locally best in its conflict graph. Although it still a long time for PP, the average time for each column is relatively short. So this so-called Multi-CG model has these new key features. First it creates a set of maximal reduced networks, that is, the subnetworks cover all the trains, but no subnetwork can be covered by another. Then it allows the subnetwork to have intersections. Finally, there is at most one source node, but multiple destination nodes allowed per subnetwork.

Variables

$\textsc{src}_v = 1$ if a consist travel plan under construction starts at node v, 0 otherwise, for $v \in \delta^-(L_c^T)$. And same situation is applied to $\textsc{dst}_v$ for its end node.
$x_\ell = 1$ if the link $\ell \in L^T \cup L^W$ belongs to the path supporting any of the consist travel plans, 0 otherwise.
n_ℓ^k: amount of locomotive flow of type k through $\ell \in L^T \cup L^W$. Note that in this PP, $n_{\ell^t}^k > 0$ iff $x_{\ell^t} = 1$, but no such limit for $n_{\ell^w}^k$. These decision variables work for consist travel plan assembly for each maximum clique.
Note that the flow decision variables x_ℓ and n_ℓ^k have no path indices, for that each path is source node-disjoint and so independent.
C represents the set of conflict graph c.

Objectives

$$
\begin{aligned}
\overline{\mathrm{COST}} = \sum_{k \in K^r \cup K^c} \sum_{\ell \in \omega^-(v^{\mathrm{SINK}}) \cap L^W} n_\ell^k \\
&- \sum_{k \in K_r \cup k_c} \sum_{v \in V \setminus (V^{\mathrm{SRC}} \cup \{v^{\mathrm{SINK}}\} \cup \delta^+(L^{\mathrm{SHOP}}))} \\
&\quad u_{kv}^{(41)} \cdot \left(\sum_{\ell \in \delta-(v)} n_\ell^k - \sum_{\ell \in \delta+(v)} n_\ell^k \right) \\
&- \sum_{k \in K_r} \sum_{v \in \delta^+(L^{\mathrm{SHOP}})} u_{kv}^{(42)} \cdot \left(\sum_{\ell \in \delta-(v)} n_\ell^k - \sum_{\ell \in \delta+(v)} n_\ell^k \right) \\
&- \sum_{k \in K_c} \sum_{v \in \delta^+(L^{\mathrm{SHOP}})} u_{kv}^{(43)} \cdot \left(\sum_{\ell \in \delta-(v)} n_\ell^k - \sum_{\ell \in \delta+(v)} n_\ell^k \right) \\
&+ \sum_{k \in K_r \cup k_c} \sum_{v \in V \setminus (V^{\mathrm{SRC}} \cup \{v^{\mathrm{SINK}}\} \cup \delta^+(L^{\mathrm{SHOP}}))} u_{kv}^{(44)} \cdot \sum_{\ell \in \delta+(v)} n_\ell^k \\
&+ \sum_{v \in \delta^+(L^{\mathrm{SHOP}})} u_{k,v}^{(45)} \cdot \sum_{\ell \in \delta+(v)} n_\ell^k \\
&+ \sum_{v \in \delta^+(L^{\mathrm{SHOP}})} u_{k,v}^{(46)} \cdot \sum_{\ell \in \delta+(v)} n_\ell^k \\
&- \sum_{\ell \in L^T} u_\ell^{(47)} \cdot x_\ell. \qquad (50)
\end{aligned}
$$

Constraints

$$\sum_{v \in \delta^-(L^T \cap c)} \mathrm{SRC}_v \leq 1 \qquad c \in C \tag{51}$$

$$\sum_{v \in \delta^+(L^T \cap c)} \mathrm{DST}_v = \sum_{v \in \delta^-(L^T \cap c)} \mathrm{SRC}_v \qquad c \in C \tag{52}$$

$$\sum_{\ell \in \omega^+(v)} x_\ell - \sum_{\ell \in \omega^-(v)} x_\ell = -\mathrm{DST}_v$$

$$v \in \delta^+(L^T \cap c), c \in C \tag{53}$$

$$\sum_{\ell \in \omega^+(v)} x_\ell - \sum_{\ell \in \omega^-(v)} x_\ell = \mathrm{SRC}_v \qquad v \in \delta^-(L^T) \tag{54}$$

$$\sum_{\ell \in \omega^+(v)} x_\ell - \sum_{\ell \in \omega^-(v)} x_\ell = 0$$

$$v \in V \cap c \setminus \left(\delta^+(L^T) \cup \delta^-(L^T) \cup V^{\mathrm{SRC}} \cup v^{sink}\right)$$

$$c \in C \tag{55}$$

$$\sum_{\ell \in \omega^-(v) \& \ell \notin L^T} x_\ell = 0 \qquad v \in V^{\mathrm{SRC}} \tag{56}$$

$$\sum_{\ell \in \omega^+(v^{sink}) \& \ell \notin L^T} x_\ell = 0 \tag{57}$$

$$x_\ell \geq \mathrm{SRC}_{\delta+(\ell)} \qquad \ell \in L^T \tag{58}$$

$$x_\ell \geq \mathrm{DST}_{\delta-(\ell)} \qquad \ell \in L^T \tag{59}$$

$$\sum_{k \in K_r} \mathrm{HP}_k \cdot n_\ell^k + \sum_{k \in K_c} \mathrm{HP}_k \cdot n_\ell^k \geq x_\ell \cdot \mathrm{HP}_t$$

$$\ell (\equiv t) \in L^T \tag{60}$$

$$\sum_{k\in K_r\cup K_c} n_\ell^k \leq M \cdot x_\ell$$

$$\ell \in L \setminus \left(\omega^+(V^{\text{SRC}}) \cup \omega^-(v^{sink})\right) \tag{61}$$

$$\sum_{k\in K_r\cup K_c} \sum_{\ell\in\omega^-(v)\cap\omega^+(v')\cap L^W} n_\ell^k \leq M \cdot \text{SRC}_v$$

$$v \in V \setminus V^{\text{SRC}}, v' \in V^{\text{SRC}} \tag{62}$$

$$\sum_{k\in K_r\cup K_c} \sum_{\ell\in\omega^-(v^{\text{SINK}})\cap\omega^+(v)\cap L^W} n_\ell^k \leq M \cdot \text{DST}_v$$

$$v \in V \setminus v^{\text{SINK}} \tag{63}$$

$$\sum_{\ell\in\omega^-(v)\cup L^W\cup L^T} n_\ell^k = \sum_{\ell\in\omega^+(v)\cup L^W\cup L^T} n_\ell^k$$

$$v \in V \setminus \left(V^{\text{SRC}} \cup v^{\text{SINK}}\right), k \in K_r \cup K_c \tag{64}$$

$$\text{consist\_size}_{\min} \leq \sum_{k\in K_r} n_\ell^k + \sum_{k\in K_c} n_\ell^k \leq \text{consist\_size}_{\max}$$

$$l \in L^T. \tag{65}$$

Constraints (51) guarantee that at most 1 source node can be selected for each conflict graph. Constraints (52) limit exactly one destination node for each source node. Constraints (53) , (54) , and (55) are the flow conservation constraints for train source nodes, train destination nodes, and other nodes (except source nodes and dummy sink node), respectively. Constraints (56) & (57) avoid the path to start from a waiting link of the station source node, or reach the dummy sink node, so to prohibit to create a path without any train link. Constraints (58) guarantee the if a source node is selected, the train starts from it must be selected, for that train links are node-disjoint. Constraints (59) work similar to (58), for the destination nodes. This flow conservation set from (53) to (59) , generate the paths which require for the locomotives assigned for the next step. Constraints (60) satisfy the power requirement of selected trains. Constraints (61), (62) and (63) avoid to assigned locomotives to the trains out of the paths selected. Constraints (64) are normal flow conservation constraints. Constraints (65) set the minimum and maximum number of locomotives for each consist.

# Trains	Model	LP Obj.	Total Time	ILP Obj.	# Columns		Round	Loc. Req.	GAP(%)
					Generated	Selected			
7-day 862-train	LAP-SCG	630,182	20h19m36s	637,980	2,207	516	4	962	1.24
	LAP-SCG+		11h03m21s						
	LAP-MCG	630,898	1h36m02s	635,500	1,543	506	155	961	0.73
+100 trains	LAP-SCG	675,010	23h22m53s	683,280	2,387	534	4	1,014	1.23
	LAP-SCG+		13h32m22s						
	LAP-MCG	674,918	2h13m05s	681,360	2,027	521	203	1,006	0.95
+200 trains	LAP-SCG	729,404	26h34m12s	737,320	2,507	541	4	1,129	1.09
	LAP-SCG+		14h38m55s						
	LAP-MCG	730,585	3h08m55s	735,520	2,325	533	233	1,109	0.68
14-day 1,750-train	LAP-SCG	1,057,813	44h41m53s	1,077,100	1,835	1,294	4	1,289	1.82
	LAP-SCG+		26h29m31s						
	LAP-MCG	1,057,949	11h18m55s	1,071,300	1,263	1,284	127	1,290	1.26
+100 trains	LAP-SCG	1,113,992	50h58m21s	1,126,780	1,654	1,355	4	1,350	1.15
	LAP-SCG+		28h19m45s						
	LAP-MCG	1,113,926	12h50m21s	1,129,860	1,544	1,355	155	1,350	1.43
+200 trains	LAP-SCG	1,165,227	56h19m41s	1,178,880	1,853	1,345	4	1,464	1.17
	LAP-SCG+		31h17m56s						
	LAP-MCG	1,164,679	21h27m17s	1,185,140	1,727	1,317	173	1,481	1.76

Table 3: Computational Comparison of the Different CG Model/Algorithm

3.6 Numerical Results

The primary objective of this study is to provide a new CG process for optimization model for real world locomotive assignment problem. We apply the train schedule and network structure from CPR, except for the larger data sets, that with a set of artificial trains.

3.6.1 Data Instances

There are 9 different types of locomotives, which are doubled by the sub-types of regular and critical ones.

The data set comes from two time period: one week and two weeks which CPR prefers for that a normal cross-continental train takes more than 5 day, and a round trip within 2 weeks. The artificial 100 and 200 trains are added to the basic data sets of 862-train schedule in 7 days and 1,750-train schedule in 14 days, so for more and larger data sets for test.

All test scenarios are solved by CPlex 12.6.1 in a computer with 40-cores, 1TB memory. To get the ϵ-optimal solution of ILP from optimal LP solution, we apply a simple rounding off procedure. In the result below, Table 3 shows that the gap between these two values are very small and our ϵ-optimal ILP solution has good quality.

# Trains	Model	Train String Size		Avg. Loc. Active Time	Loc. Use Rate (%)	In-Consist Idle Time		HP Var. per Train	Loc. Fleet Size
		Avg.	Max.			Average	Maximum		
7-day 862-train	Original	1.65	10	3d05h10m	45.9	9h38m	3d18h00m	199.3	962
	Multi-CG	1.70	11	3d04h29m	45.5	9h31m	3d14h53m	191.4	961
+100 tr	Original	1.77	8	3d03h53m	45.2	9h52m	3d16h55m	178.4	1,014
	Multi-CG	1.85	10	3d04h34m	45.6	8h58m	3d14h53m	187.2	1,006
+200 tr	Original	1.92	10	3d02h34m	44.4	9h47m	3d20h43m	174.8	1,129
	Multi-CG	1.99	11	3d04h38m	45.6	9h22m	3d14h53m	180.6	1,109
14-day 1,750-train	Original	1.35	13	6d10h03m	45.9	18h53m	8d03h19m	106.7	1,289
	Multi-CG	1.36	16	6d15h20m	47.4	18h27m	8d03h19m	100.9	1,290
+100 tr	Original	1.37	18	6d12h57m	47.3	16h25m	7d04h10m	92.9	1,350
	Multi-CG	1.37	20	6d16h38m	47.8	14h47m	6d16h33m	94.3	1,350
+200 tr	Original	1.45	18	6d12h04m	46.4	18h07m	8d03h19m	93.1	1,464
	Multi-CG	1.48	20	6d15h24m	47.7	17h05m	8d03h19m	94.6	1,481

Table 4: Characteristics of LAP Solutions

3.6.2 Computational Comparison of the Different CG Models

Table 3 provides the solution of these scenarios from three LAP models: original (marked as LAP-SCG in the table), PP with conflict graph (marked as LAP-SCG+), and multiple column generated (marked as LAP-MCG, 10 columns generated per call of PP).

The columns in Table 3 provides the data set characteristics, the objective value and total computational time for LP and ILP, the number of consist travel plans columns that PP generated, and selected by the final solutions. The column "rounds " shows how many cycles to check all the possible origin nodes. The second last column shows the total number of locomotives needed. The last column provides the solution quality, i.e., the gap between the optimal solution of LP and the ϵ-optimal solution of ILP.

Table 3 shows that pre-process the initial network reduce computational time by half. The multi-CG process reduce the total by another half, without touching the quality of final solution.

3.6.3 Characteristics of LAP Solutions

Table 4 lists several aspects of results, for original and multi-CG LAP models. LAP with conflict graph is excluded because its result is the same as the original one. The difference of the results of two models are basically the default process of LP to ILP solution by CPlex. The locomotive usage rates are less than 50%, because first, we set 20% of locomotives to undergo the maintenance shop process, and second, the compromise of locomotive usage and the locomotive relocation process by deadheading.

3.6.4 Analysis of Multi-CG Architecture

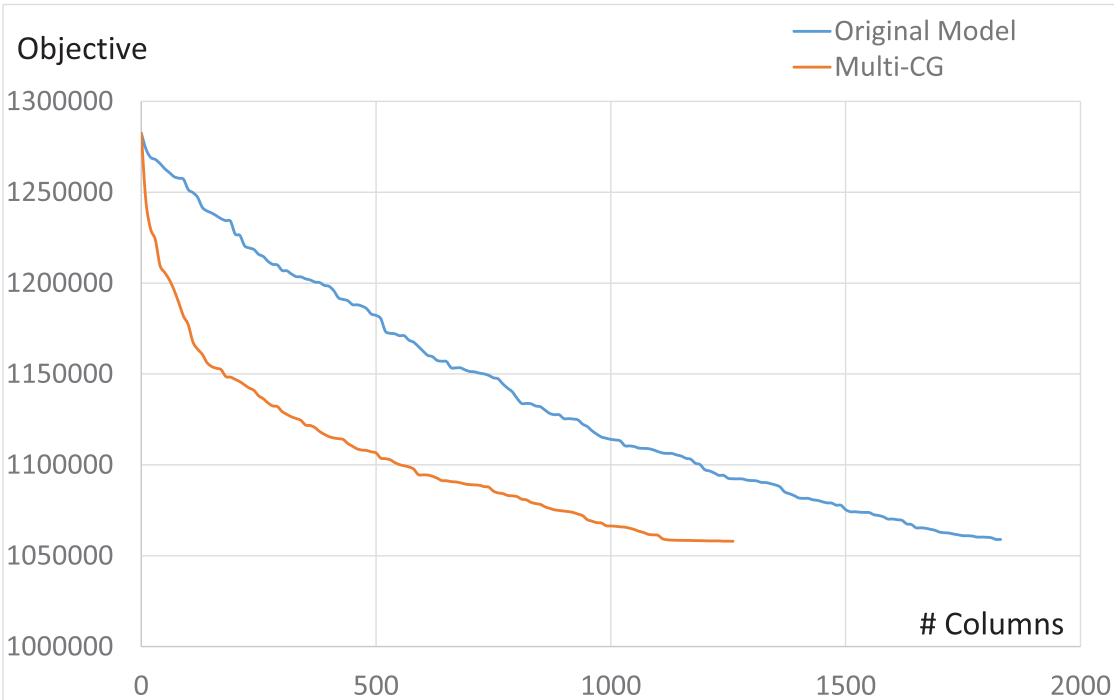


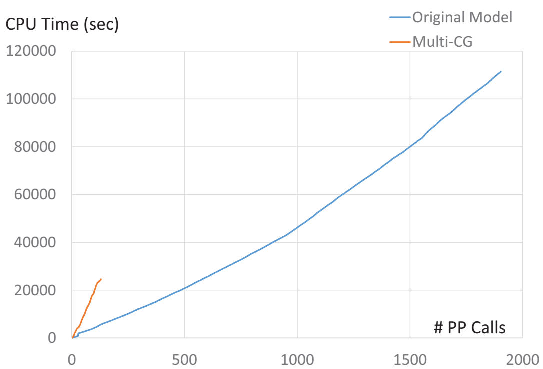
Figure 13: Comparison: Converge Rate

The computational time saving comes from two aspects. The first one is the higher convergence rate per each PP call. Comparing to the original model which fix the source node to each PP, the PP of Multi-CG model search globally form the original node of entire network, so to generate the global optimal column to feed back to RMP. Figure 13 shows the different converge rate of both model, by the curves of objective value over the number of columns added to RMP. The Multi-CG model uses less columns to reach the optimal solution for the global optimal columns from PP.

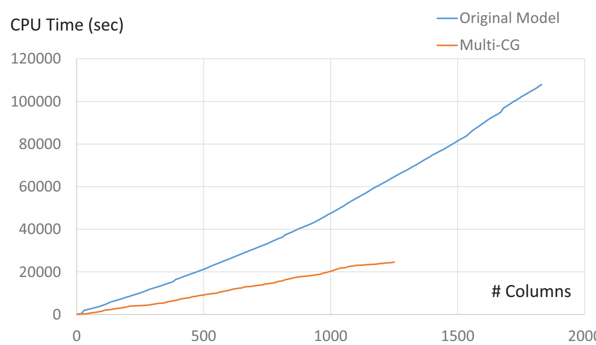
Secondly, the Multi-CG model uses less average time per column showed in Figure 14(b), for that its PP generate 10 columns, even in Figure 14(a) each PP takes more time.

3.7 Conclusion

We propose in the paper a new CG process with multiple columns generated by each pricing problem call. Applying the new CG process to the original LAP can save 60-93% of computational time without reducing the quality of final solution. This new multi-CG process may be applied to general network flow models which has potential network decomposition into several overlapped sub-networks.



(a) CPU time over PP Call



(b) CPU Time Over Column

Figure 14: Comparison: CPU Time Analysis

Chapter 4

Minimizing Fuel Consumption for Freight Trains

H. Tian and B. Jaumard. submitted for publication, 2017.

4.1 Introduction

Freight train transportation is more important in logistics, for the energy efficiency by the low friction over the rail. Today to reduce the energy consumption additionally is a key topic in railway industry for lower emission for global warming and also fuel cost saving. The optimization of freight train energy/fuel consumption is basically influenced by two factors: train running time & speed by railway scheduling and the driving strategy. In fact, the train fuel is consumed by its consist which is a group of locomotives to tract the train. This paper focuses on the first part, to optimize freight train fuel consumption by an optimization model for locomotive assignment with train (re)scheduling.

Based on our previous works Jaumard *et al.* [33][36][34][32], this paper integrates the train fuel consumption model to previous LAP model, for reducing the energy usage for the locomotive fleet on given freight train schedule. Further more, we notice that if some trains can make adjustment to their scheduling, e.g., late departure, early arrival, or shift without influencing the goods transportation much, the request of locomotives and the fuel consumption should be reduced greatly. So we investigate the combination of LAP model, fuel consumption model, and the train (re)scheduling feature together so as to reduce the total costfuel consumption in railway operation.

This paper is designed as follows. In Section 4.2, we take a review of the literatures for LAP model with and without time window, and fuel consumption model to evaluate train energy consumption. Section 4.3 generally describe the enhanced LAP model. Section 4.4

gives the details of enhanced LAP model with column generation (CG) decomposition over relaxation of time window. In Section 4.5, the fuel consumption model is discussed and integrated to LAP model. and in Section 4.6, we provides the numerical results and their analysis.

4.2 Literature Review

Here we lists the review of the works on exact mathematical models in locomotive assignment problem area. The works in heuristics can be checked the survey of Piu *et al.* [45]. Here we do not repeat the basic concepts and rule in LAP and railway industry which was described in our previous papers Jaumard *et al.* [33][36][34][32].

Ziarati *et al.* [55] solve LAP as a classical integer multi-commodity network flow problem with the nonlinear part for maintenance constraints, i.e., maintain the locomotive regularly in shop site. They use column generation (CG) format to solve the model. Ziarati *et al.* decompose the full problem into sub-problems of 2 or 3 days for feasible solution to avoid the scalability issue. Rouillon *et al.* [48] improve Ziarati *et al.* [55] with different branching methods and search strategies by a branch-and-price algorithm. Ahuja *et al.* [9] developed a MILP model of LAP also as a locomotive flow model that similar to the one of Ziarati *et al.* [55], but assuming that the train schedule and the locomotive assignment plan is cyclic every week. However, they do not consider the maintenance process, neither for the consist busting issue.

The models of Ziarati *et al.* [55] and Ahuja *et al.* [10] can solve certain LAP problem but with the high rate of consist busting (the medium rate is 50%) which means the solution intends to break the consist and assign the locomotives to different trains. Then their MIP models have severe scalability issues to solve the problems that the railway industry really needed. So Vaidyanathan *et al.* [54] create a consist-based LAP model. A pre-processing algorithm is used to generate configurations of consists. The model assigns consists instead of assigning locomotives. The consist-based model requires greater numbers of configuration types of consist so for longer computational time with larger data instance. In addition Vaidyanathan *et al.* [54] do not consider the maintenance constraint.

Other research works are available for problems similar to LAP. Cordeau *et al.* [22] solved the locomotive and car assignment in passenger transportation, with an exact model based on the Benders decomposition approach without maintenance constraint. Fügenschuh *et al.* [26] focused on locomotive and car cycle scheduling problem with time window without maintenance but allow the train delay within given time window. Cacchiani *et al.* [17] provide two ILP models on the train unit assignment problem in passenger transportation by combination of passenger cars and supported locomotive for certain trips, without the maintenance constraint. One model is solved by linear programming (LP) relaxation based heuristic, the other by Lagrangian relaxation.

Our previous work Jaumard *et al.* [33][36][34][32] propose a consist travel plan (previously called train string) based optimization model which considers locomotive maintenance constraint and consist busting issei. Using our multi-column generation algorithm, the computational time is greatly reduced to several hours to solve the largest 2-week train schedule with 1,750 scheduled trains and 9 types of locomotives over entire Canada Pacific Railway network.

To calculate train fuel consumption, there are several models available based on the Davis Formula in 1920s: Radford [47], Ghoseiri *et al.* [28], Condie [21], Koç *et al.* [39] [40] and also in AREMA manual [11]. We provide the details in Section 4.5.

4.3 LAP Model

We next develop the LAP optimization model we propose for the locomotive assignment, with the consideration of the optimization of total fuel consumption. In order to alleviate the presentation, we describe it without the legacy trains.

4.3.1 Notations

T is the set of trains t in given schedule.
$C_t \in C$ is set of trains other than $t \in T$, that there is a time constraint between t and $t' \in C_t$ for their departure time. $c_{t,t'}$ represents the value of this time, usually negative.
$D_t \in D$ is set of trains other than $t \in T$, that there is a time constraint between t and $t' \in D_t$ for their arrival time. $d_{t,t'}$ represents the value of this time, usually positive.
S is the set of consist travel plans, where a consist travel plan $s \in S$ defines a sequence of trains (original or rescheduled train occurrences) led by the same locomotive consist. Note that a consist travel plan also store the departure/arrival time of rescheduled train.
S_v^+ is the set of consist travel plans departing from node v, S_v^- is the set of consist travel plans arrive at node v K is the set of locomotives, indexed by k, which represents a certain locomotive. Each locomotive k has several parameters, e.g., the horsepower HP_k, and the subtype of regular/critical ($k_r \cup k_c \in k$).
n_k^s represents number of locomotives of type k assigned to the consist travel plan $s \in S$.
$d_t^s = 1$ if train $t \in T$ belongs to consist travel plan s, 0 otherwise.
$SrcTime_t^s$ the original or rescheduled departure time of $t \in T$ belongs to consist travel plan

s. If train $t \in T$ does not belongs to consist travel plan s, $SrcTime_t^s = 0$.
$DstTime_t^s$ the original or rescheduled arrival time of $t \in T$ belongs to consist travel plan s. If train $t \in T$ does not belongs to consist travel plan s, $DstTime_t^s = 0$.
Note that d_t^s, $SrcTime_t^s$ and $DstTime_t^s$ are not decision variables, but attributes of consist travel plan s.
Note that the rescheduled consist travel plan r equals to consist travel plan s so all s above can be replaced by r.
$n_{k,v}^{\text{SPARE}}$ = number of spare locomotives of type k in source node $v \in V^{\text{SRC}}$
$\text{CAP}(\ell^{\text{SHOP}})$ = upper bound of critical locomotives that can be maintained in shop link $\ell^{\text{SHOP}} \in L^{\text{SHOP}}$.

In the multi-commodity graph $G = (V, L)$, we designate by $\omega(v)$ (resp. $\omega(V')$ with $V' \subseteq V$) the set of incident links to v (resp. to a node of V'). In addition, $\omega^+(v)$ (resp. $\omega^-(v)$) denotes the set of incident outgoing (resp. incoming) links of v. For a given link ℓ, $\delta^+(\ell)$ denotes the destination endpoint of ℓ, and $\delta^+(L'), L'' \subseteq L$ denotes the set of destination endpoints of the links of L'. Similarly, $\delta^-(\ell)$ and $\delta^-(L'), L'' \subseteq L$ denote the origin endpoint(s) of ℓ and of the links of L', respectively.

4.3.2 Variables.

We use two sets of variables:
$z_s = 1$ if string s is selected, 0 otherwise.
$x_{k\ell}^{\text{LOCO}}$ represents the number of locomotive k goes through link $\ell \in L^W \cup L^D \cup L^{\text{SHOP}}$.

4.3.3 Objective

The primary objective is to minimize the total fuel consumption of the locomotives in operation, then for the number of consist busting and the size of the locomotive fleet. So we modified the previous objective of LAP, converting the cost of consist busting and locomotive lease to certain fuel requirement that marked as the penalties for each segment.

(i) penalty in fuel requirement of the number of total locomotives in operation; *(ii)* penalty in fuel requirement of the number of consist travel plans ; *(iii)* the fuel consumption of each consist travel plan.

$$
\begin{aligned}
\min \quad & \sum_{\ell \in \omega^-(v^{\text{SINK}})} \sum_{k \in K} \text{PENAL}^{\text{LOCO}} \cdot x_{k\ell}^{\text{LOCO}} + \sum_{s \in S} \sum_{k \in K} \text{PENAL}^{consist\_bust} \cdot n_k^s \, z_s \\
& + \sum_{s \in S} \text{FuelCost}_s \, z_s
\end{aligned}
$$

(On this stage, we do not include of the fuel consumption model yet, so the objective is simplified as below.)

$$
\text{Obj\#1:} \qquad \min \quad \sum_{\ell \in \omega^-(v^{\text{SINK}})} \sum_{k \in K} x_{k\ell}^{\text{LOCO}} + \sum_{s \in S} \sum_{k \in K} z_s \tag{66}
$$

$\text{PENAL}^{\text{LOCO}}$ represents the lease and operation cost of one locomotive in fuel requirement. $\text{PENAL}^{consist\_bust}$ represents the consist busting cost for each locomotive in fuel requirement.

4.3.4 Constraints

$$\sum_{\ell \in \omega^-(v^{\text{SINK}})} x^{\text{LOCO}}_{k\ell} \leq n_k \qquad k \in K. \tag{67}$$

$$\begin{aligned}&\sum_{s \in S^+_v} n^s_k\, z_s + \sum_{\ell \in \omega^+(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x^{\text{LOCO}}_{k\ell} \\ &\quad = \sum_{s \in S^-_v} n^s_k\, z_s + \sum_{\ell^w \in \omega^-(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x^{\text{LOCO}}_{k\ell} \\ &\qquad v \in V \setminus \left(V^{\text{SRC}} \cup \{v^{\text{SINK}}\} \cup \delta^+(L^{\text{SHOP}})\right), k \in K_r \cup K_c\end{aligned} \tag{68}$$

$$\begin{aligned}&\sum_{s \in S^+_v} n^s_{k_r}\, z_s + \sum_{\ell \in \omega^+(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x^{\text{LOCO}}_{k_r\ell} \\ &\quad = \sum_{s \in S^-_v} n^s_{k_r}\, z_s + \sum_{\ell \in \omega^-(v) \,\cap\, L^{\text{SHOP}}} x^{\text{LOCO}}_{k_c\ell} + \sum_{\ell \in \omega^-(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x^{\text{LOCO}}_{k_r\ell} \\ &\qquad v \in \delta^+(L^{\text{SHOP}}), k = \{k_r, k_c\} \in K\end{aligned} \tag{69}$$

$$\begin{aligned}&\sum_{s \in S^+_v} n^s_k z_s + \sum_{\ell \in \omega^+(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x^{\text{LOCO}}_{k\ell} = \sum_{s \in S^-_v} n^s_k z_s + \sum_{\ell \in \omega^-(v) \,\cap\, (L^{\text{WAIT}} \cup L^D)} x^{\text{LOCO}}_{k\ell} \\ &\qquad v \in \delta^+(L^{\text{SHOP}}), k \in K_c.\end{aligned} \tag{70}$$

$$\sum_{s \in S} d^s_t \cdot z_s = 1 \qquad t \in T \tag{71}$$

$$\sum_{k \in K} x^{\text{LOCO}}_{k_c \ell^{\text{SHOP}}} \leq \text{CAP}(\ell^{\text{SHOP}}) \qquad \ell^{\text{SHOP}} \in L^{\text{SHOP}} \tag{72}$$

$$\sum_{s \in S} SrcTime^s_t \cdot z_s - \sum_{s \in S} SrcTime^s_{t'} \cdot z_s \leq c_{t,t'} \qquad t \in T, t' \in C_t \tag{73}$$

$$\sum_{s \in S} DstTime^s_t \cdot z_s - \sum_{s \in S} DstTime^s_{t'} \cdot z_s \geq d_{t,t'} \qquad t \in T, t' \in D_t \tag{74}$$

Constraints (67) guarantee that the solution will not exceed the total locomotives available for each type. Constraints (68), (69) and(70) are the flow conservation constraints for normal nodes and shop end nodes, excluding the source and dummy sink nodes. Note that critical locomotives are relabelled as regular after completing the maintenance process at a shop node. This relabelling will be taken care thanks to special flow conservation constraints at the shop end nodes in the proposed LAP model. Constraints (71) guarantee that for each original train, there should be one and **exactly** one consist travel plan in the locomotive assignment that covers it or one of its corresponding train occurrences. Constraints (72) take effective to limit the number of critical locomotives at the same time for maintenance within the shop capacity. Constraints (73) and (74) are the conflict-free source/dsetination time constraints, which will be discussed in Section 4.4.2.

4.4 Pre-processing for Conflict-free Consist Travel Plans with Rescheduled Trains

RMP deals with a set of consist travel plans, where a consist travel plan is defined by a path, such that the first and last links are associated with (rescheduled) train links. Difference with the past (add a ref. later) is that we allow trains to have a different departure/arrival time. While we limit the number of consist travel plans, we still allow all possible starting nodes for the consist travel plans, i.e., among the origin nodes of the trains. Note that that set includes not only the original source nodes of the trains (as originally scheduled), but also, all the new origin nodes of the rescheduled trains. For each train, only one origin node will be selected.

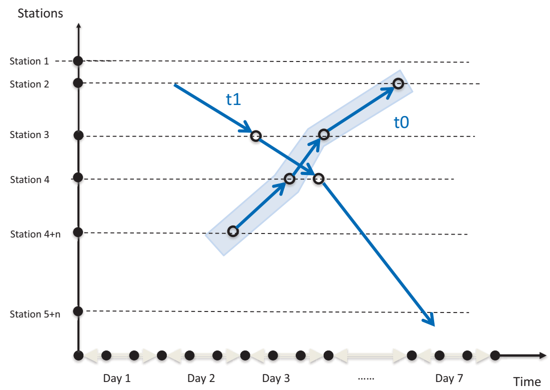
We have alternative choice to relax the train schedule for LAP problem to reduce the energy consumption. For the pricing problem, we allow to reschedule train(s) for better usage of locomotives. That means to reduce the total number of locomotives, avoid some of the consist busting and deadheading operations, and to adjust train speed to save the total fuel/energy consumption within the give time period.

4.4.1 Train Rescheduling Allowed within LAP

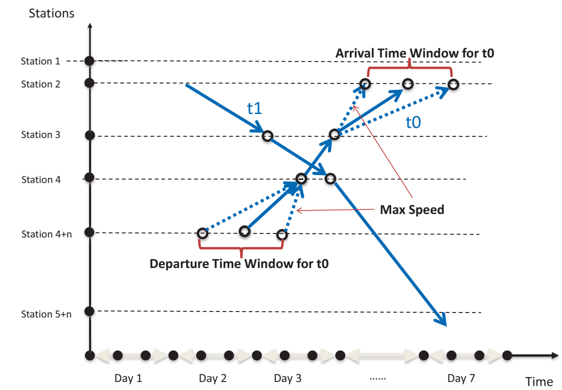
We set some limitations for train rescheduling in LAP. Firstly, the time length of a rescheduled train should not exceed ±20% of the original length. Secondly, the adjustment of departure/arrival time should not exceed 6 hours.

4.4.2 Avoidance of Train Rescheduling Conflict

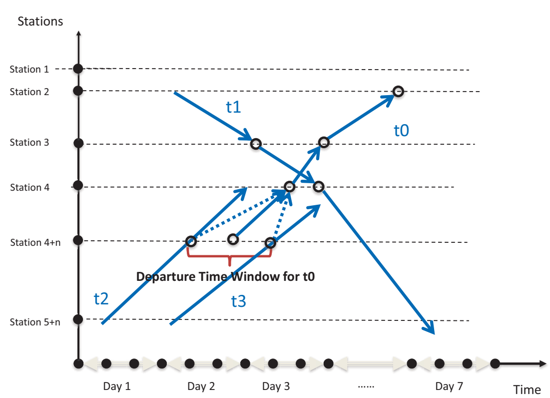
For LAP problem, a train schedule is given without conflict. But with train rescheduling the new model may cause train conflicts with new deaprture/arrival times. We propose an algorithm to avoid potential train conflict during the pre-processing of the given schedule. Firstly, for train rescheduling, all existing cross point should not be changed in the time-space network of train schedule. As Figure 15(a) shows, we plan to reschedule train t_0 which meets train t_1 between station 3 & 4. Train t_0 can be divided into three parts, and the middle part, between station 3 & 4, should be kept so that the cross point with t_0 is untouched. Then, as Figure 15(b) shows, from the maximum speed limit of the source station to station 4, we get the latest departure time allowed for t_0 rescheduling. The earliest departure time should be set as a maximum allow time for a fixed time, e.g., 6 hours, or a ration of the time length of t_0, e.g., 15%. The similar process will be applied to the arrival time rescheduling. If there are multiple cross points to t_0, we take the earliest one for departure time window, and the



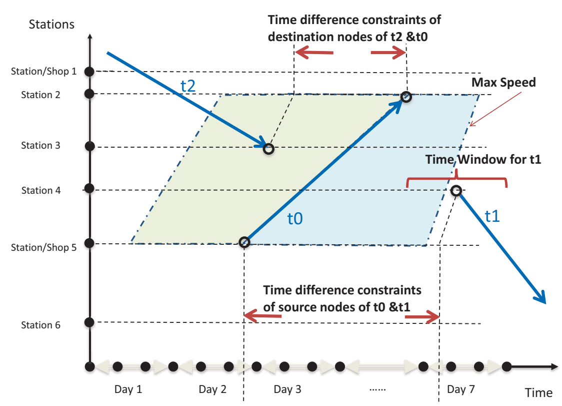
(a) Original trains and cross point



(b) Reschedule train without touch cross point



(c) Reschedule train without new cross point



(d) Reschedule train without new cross point

Figure 15: Reschedule Train without Conflict

latest one for arrival time window. With this pre-process applied to each train, we guarantee that the train rescheduling will not touch the cross points with the given time windows.

In addition, rescheduled train may cause new crossing with other trains which are not meet with the original train. So after the first step, there is another procedure to check within the time windows of train t_0, if there are trains depart or arrive without crossing point but have the potential meet with rescheduled train. Figure 15(c) shows the for the departure time of train t_0, it can not exceed the time of another train that passing through the same station, e.g., trains t_2 and t_3. The same condition is applied to the arrival time window.

There is a more complex situation, as Figure 15(d) shows. The colored areas represent the possible positions for t_0 rescheduling. If another train have a departure/arrival node inside, or even its possible time window has an overlap, e.g., t_1 (and also t_2) in the figure, t_1 may meet t_0 if one or both rescheduled. In order to avoid this kind of potential additional cross point, a limit on departure/arrival time windows is not enough. A time constraint is needed to set a minimal time period between the departure/arrival nodes of these two trains, e.g., t_0 and t_1, to guarantee that with the maximum speed, t_0 can not arrive the same station before t_1's departure. This set of constraints will be applied to RMP.

4.4.3 Multiple Consist Travel Plans Generator

Our pricing problem inhabit the multiple consist travel plans generation feature from our previous work Jaumard *et al.* [33]. Now each train has its reduced network which includes all the train links and waiting links it can connected to. And the conflict graph is the reduced network that can not be a entire subset of any other reduced network. For each conflict graph the pricing problem can generate at most one consist travel plan.

4.4.4 Notations and Variables

A consist travel plan is defined by a path, such that the first and last links are associated with (rescheduled) train links.

In order for the PP to output a given number of consist travel plans, i.e., a set P of paths, the PP must compute $|P|$ train link disjoint paths, with each path being indexed by p.

$\alpha^{\text{SRC}}_{\ell,v,p} = 1$ if train link $\ell \in L^T$ is selected by consist travel plan $p \in P$, with a train departure at node v, 0 otherwise. Note that, if $v \neq \delta^+(\ell)$, then departure of train ℓ is modified wrt. the original schedule.

$\alpha^{\text{DST}}_{\ell,v,p} = 1$ if train link $\ell \in L^T$ is selected by consist travel plan $p \in P$, with a train arrive at node v, 0 otherwise. Note that, if $v \neq \delta^-(\ell)$, then arrival of train ℓ is modified wrt. the original schedule.

Note that even the train is not rescheduled, it still has $\alpha^{\text{SRC}}_{\ell,v,p} = 1$ and $\alpha^{\text{DST}}_{\ell,v,p} = 1$ at its source/destination node. In addition, we do not allow a train to connect directly to the station source node or the dummy sink node.

$p \in P$ represents the serial number assigned to a consist travel plan path.

$x_{\ell,p} = 1$ if wait link $\ell \in L^W$ belongs to consist travel plan path $p \in P$, 0 otherwise.

$\text{SRC}_{v,p} = 1$ if consist travel plan path p under construction starts at node v, 0 otherwise, for $v \in \delta^-(L^T_c)$.

$\text{DST}_{v,p} = 1$ if consist travel plan path p under construction ends at node v, 0 otherwise, for $v \in \delta^+(L^T_c)$.

$n_{k,p}$: amount of locomotive flow of type k through consist travel plan path $p \in P$.

$n^{\text{SRC}}_{k,v,p}$: amount of locomotive flow of type k from node v which is the source node of a consist travel plan path $p \in P$.

$n^{\text{DST}}_{k,v,p}$: amount of locomotive flow of type k to node v which is the destination node of a consist travel plan path $p \in P$. Note these two decision variables are work for the objective only.

d_ℓ represents the running distance of a train link.

Each train link $\ell \in L^T$ has two set of nodes: V_ℓ^{SRC} for all available/connectable nodes within ℓ's source area, and V_ℓ^{DST} within its destination area.
Correspondingly, Each node $v \in V$ has two set of train links $L_v^{T,\text{SRC}}$ for all train links that can be connected or rescheduled to it as the source node, and $L_v^{T,\text{DST}}$ for those by the destination node.

4.4.5 Objective: Reduced Cost of the z_s Variables

The objective of the pricing problem is the so-called reduced cost (if not familiar with linear programming concepts, the reader is referred to, e.g., Chvátal *et al.* [20]) of the z_s variables. The s index in the master problem is replaced by consist travel plan path index p.

Firstly we set the reduced cost segment $\overline{\text{COST}}_0$ as the fix part that will not affected by the objective of master problem:

$$
\begin{aligned}
&\overline{\text{COST}}_0 = \\
&\sum_{k \in K_r \cup k_c} \left(\sum_{v \in V \setminus (V^{\text{SRC}} \cup \{v^{\text{SINK}}\} \cup \delta^+(L^{\text{SHOP}}))} u_{kv}^{(68)} \cdot \sum_{p \in P} \left(n_{k,v,p}^{\text{SRC}} - n_{k,v,p}^{\text{DST}} \right) + \sum_{v \in \delta^+(L^{\text{SHOP}})} u_{kv}^{(69)} \cdot \sum_{p \in P} \left(n_{k,v,p}^{\text{SRC}} - n_{k,v,p}^{\text{DST}} \right) \right) \\
&\quad + \sum_{k \in K_c} \sum_{v \in \delta^+(L^{\text{SHOP}})} u_{kv}^{(70)} \cdot \sum_{p \in P} \left(n_{k,v,p}^{\text{SRC}} - n_{k,v,p}^{\text{DST}} \right) + \sum_{\ell \in L^T} u_\ell^{(71)} \cdot \left(\sum_{p \in P} \sum_{v \in V_\ell^{\text{SRC}}} \alpha_{\ell,v,p}^{\text{SRC}} \right) \\
&\quad - \sum_{\ell \in L^T} \sum_{\ell' \in C_\ell} (u_{\ell,\ell'}^{(73)} \cdot (\sum_{p \in P} \sum_{v \in V_\ell^{\text{SRC}}} \alpha_{\ell,v,p}^{\text{SRC}} - \sum_{p \in P} \sum_{v \in V_{\ell'}^{\text{SRC}}} \alpha_{\ell',v,p}^{\text{SRC}})) \\
&\quad + \sum_{\ell \in L^T} \sum_{\ell' \in D_\ell} (u_{\ell,\ell'}^{(74)} \cdot (\sum_{p \in P} \sum_{v \in V_\ell^{\text{SRC}}} \alpha_{\ell,v,p}^{\text{SRC}} - \sum_{p \in P} \sum_{v \in V_{\ell'}^{\text{SRC}}} \alpha_{\ell',v,p}^{\text{SRC}})).
\end{aligned}
$$

And so now the reduced cost, i.e., objective of pricing problem is:

$$
\overline{\text{COST}}_1 = \text{SIZE}(P) - \overline{\text{COST}}_0. \tag{75}
$$

Note the reduced cost is the sum of the multiple consist travel plans generated by the pricing problem. However, there may be some of them with positive reduced costs. So in the post process of PP, we calculate the reduced cost for each consist travel plan of the solution and only add those with negative values. And also the stopping condition takes effect when no column with negative reduced cost can be found by PP.

4.4.6 Constraints

$$\sum_{p\in P}\sum_{v\in V_\ell^{\text{SRC}}} \alpha^{\text{SRC}}_{\ell,v,p} \leq 1 \qquad \ell \in L^T \tag{76}$$

$$\sum_{p\in P}\sum_{v\in V_\ell^{\text{DST}}} \alpha^{\text{DST}}_{\ell,v,p} \leq 1 \qquad \ell \in L^T \tag{77}$$

$$\sum_{v\in V_\ell^{\text{SRC}}} \alpha^{\text{SRC}}_{\ell,v,p} = \sum_{v\in V_\ell^{\text{DST}}} \alpha^{\text{DST}}_{\ell,v,p} \qquad p\in P, \ell \in L^T \tag{78}$$

$$\sum_{p\in P} \alpha^{\text{SRC}}_{\ell,v,p} = \sum_{p\in P} \alpha^{\text{DST}}_{\ell,v,p} = 0 \qquad \ell \in L^T, v\in V^{\text{SRC}} \cup \{v^{sink}\} \tag{79}$$

$$\sum_{p\in P}\sum_{v\in V\setminus V_\ell^{\text{SRC}}} \alpha^{\text{SRC}}_{\ell,v,p} = \sum_{p\in P}\sum_{v\in V\setminus V_\ell^{\text{DST}}} \alpha^{\text{DST}}_{\ell,v,p} = 0 \qquad \ell \in L^T, \tag{80}$$

$$\sum_{p\in P}\sum_{v\in V_\ell^{\text{SRC}}}\sum_{\ell\in L^T} \alpha^{\text{SRC}}_{\ell,v,p} \geq 1 \tag{81}$$

$$\sum_{p\in P}\sum_{v\in c} \text{SRC}_{v,p} \leq 1 \qquad c\in C \tag{82}$$

$$\sum_{p\in P}\sum_{v\in V} \text{SRC}_{v,p} \geq 1 \tag{83}$$

$$\sum_{v\in V} \text{DST}_{v,p} = \sum_{v\in V} \text{SRC}_{v,p} \leq 1 \qquad p\in P \tag{84}$$

$$\sum_{p\in P} \text{SRC}_{v,p} = \sum_{p\in P} \text{DST}_{v,p} = 0 \qquad v\in V^{\text{SRC}} \cup \{v^{sink}\} \tag{85}$$

$$\text{SRC}_{v,p} + \text{DST}_{v,p} \leq 1 \qquad p\in P, v\in V \tag{86}$$

$$\text{SRC}_{v,p} \leq \sum_{\ell\in L^T} \alpha^{\text{SRC}}_{\ell,v,p} \qquad p\in P, v\in \bigcup_{\ell\in L^T} V_\ell^{\text{SRC}} \tag{87}$$

$$\text{DST}_{v,p} \leq \sum_{\ell\in L^T} \alpha^{\text{DST}}_{\ell,v,p} \qquad p\in P, v\in \bigcup_{\ell\in L^T} V_\ell^{\text{DST}} \tag{88}$$

$$\sum_{\ell\in L^T} \alpha^{\text{SRC}}_{\ell,v,p} + \sum_{\ell'\in L^T} \alpha^{\text{DST}}_{\ell',v,p} \leq 1 \qquad p\in P, v\in V \tag{89}$$

$$\left(\sum_{\ell\in L_v^{T,\text{SRC}}} \alpha^{\text{SRC}}_{\ell,v,p} + \sum_{\ell\in\omega^+(v)\cap L^W} x_{\ell,p}\right) - \left(\sum_{\ell\in L_v^{T,\text{DST}}} \alpha^{\text{DST}}_{\ell,v,p} + \sum_{\ell\in\omega^-(v)\cap L^W} x_{\ell,p}\right)$$
$$= \text{SRC}_v - \text{DST}_v \qquad p\in P, v\in V\setminus \left(V^{\text{SRC}} \cup \{v^{sink}\}\right) \tag{90}$$

$$\sum_{\ell \in L_v^{T,\text{DST}}} \alpha_{\ell,v,p}^{\text{DST}} + \sum_{\ell \in \omega^-(v) \cap L^W} x_{\ell,p} + \text{SRC}_v \leq 1$$

$$p \in P, v \in V \setminus \left(V^{\text{SRC}} \cup \{v^{sink}\}\right) \tag{91}$$

$$\sum_{k \in K_r \cup K_c} \text{HP}_k \cdot n_{k,p} \geq \sum_{v \in V_\ell^{\text{SRC}}} \alpha_{\ell,v,p}^{\text{SRC}} \cdot \text{HP}_t \qquad p \in P, \ell(\equiv t) \in L^T \tag{92}$$

$$\sum_{k \in K_r \cup K_c} n_{k,p} \geq 2 \qquad p \in P, \ell(\equiv t) \in L^T \tag{93}$$

$$n_{k,v,p}^{\text{SRC}} \leq M * \text{SRC}_{v,p} \qquad v \in V, k \in K, p \in P \tag{94}$$

$$n_{k,v,p}^{\text{SRC}} \leq n_{k,p} \qquad v \in V, k \in K, p \in P \tag{95}$$

$$n_{k,v,p}^{\text{SRC}} \geq n_{k,p} - M * (1 - \text{SRC}_{v,p}) \qquad v \in V, k \in K, p \in P \tag{96}$$

$$n_{k,v,p}^{\text{DST}} \leq M * \text{DST}_{v,p} \qquad v \in V, k \in K, p \in P \tag{97}$$

$$n_{k,v,p}^{\text{DST}} \leq n_{k,p} \qquad v \in V, k \in K, p \in P \tag{98}$$

$$n_{k,v,p}^{\text{DST}} \geq n_{k,p} - M * (1 - \text{DST}_{v,p}) \qquad v \in V, k \in K, p \in P \tag{99}$$

Constraints (76), (77), and (78) guarantee that for each train, it can be either not selected, or selected without change, or selected with change of departure andor arrival time.

Constraints (79) prohibit any train start from the schedule plan beginning time or end in the dummy sink node.

Constraints (80) prohibit train link ℓ to rescheduled to the node out of the given departure/arrival time window.

Constraints (81) guarantee that at least one train is selected.

Constraints (82) allow no more than 1 source node per conflict graph (the definition in Section 4.4.3).

Constraints (83) guarantee that at least one consist travel plan should be generated.

Constraints (84) guarantee that for each consist travel plan path, there should be at most one source node, at most one destination node, and should be in pair.

Constraints (85) prohibit the new consist travel plan start from the station source node, or end in the dummy sink end. (They take effect for the flow conservation constraints (90))

Constraints (86) avoid an almost-empty consist travel plan with only one node.

Constraints (87) & (88) guarantee that a consist travel plan starts from train origin and ends at the train destination only.

Constraints (89) guarantee that for any node there is at most one train connected, inbound or outbound. This constraints set prohibit at any node two train be directly connected.

Constraints (90) are the general flow conservation constraints.

Constraints (91) guarantee that for each node in the path, at most one link in, one link out.

Constraints (92) assign enough power to each consist travel plan to satisfy the power requirement of each train selected.
Constraints (93) guarantee each consist including no less than two locomotives for operation safety.
Constraints (94) to (99) are the linearlization for the locomotive flow from source node and to the destination node of each path of consist travel plan generated.

4.5 Fuel Consumption Model & Integration

4.5.1 Fuel Consumption Models

Fuel consumption model calculates for each consist travel plan $s \in S$, the fuel consumption FuelCost_s, depending on the speed and/or traveling time of all the trains in s.

Jaumard *et al.* [31] establish a freight train fuel consumption model which comes from AREMA manual [11]. From their model, the rolling resistance (used in Condie [21], Radford [47], Ghoseiri *et al.* [28]), which is affected by train speed (and weight), has been studied by Davis in 1920s as

$$R = A + B \cdot V + C \cdot V^2 \tag{100}$$

where R represents the train rolling resistance, V represents train speed, and three coefficients A, B, and C represent the resistance not affected by speed, that affected by speed, and air turbulence. Note that now we only consider the condition in flat ground and straight track and constant speed.

Radford [47] calculate the train fuel consumption rate per hour as linearly to the train horsepower:

$$F = a + b \cdot P \tag{101}$$

where F is the fuel consumption rate in gal/hr, a is the constant fuel consumption for the base load (Radford [47] uses 6 gallon/hour), P is the horsepower, and b is the proportional fuel consumption rate to the horsepower.

Based on AREMA manual [11], the relation between the tractive effort R and horsepower at a certain speed V without acceleration is:

$$R(lbs.) = \frac{P(hp) \cdot 308}{V(mph)} \tag{102}$$

in which 308 is 82% of 375 lb-miles per hour per hp, and 82% is the normal rate for modern locomotive power for traction.

Since the traveling time equals to distance d divided by speed V, finally we get the fuel consumption E in gal as

$$\begin{aligned} E(gal) &= \frac{F.d}{V} \\ &= a \cdot \frac{d}{V} + b \cdot P \cdot \frac{d}{V} \\ &= a \cdot \frac{d}{V} + b \cdot \frac{R \cdot V}{308} \cdot \frac{d}{V} \\ &= a \cdot \frac{d}{V} + b(A + BV + CV^2)\frac{d}{308} \end{aligned} \tag{103}$$

There is another fuel consumption model for general vehicle types, used by Koç *et al.* [39][40], Barth *et al.* [14], from the origin of An *et al.* [12].

$$E^h = \lambda(k^h N^h V^h d/v + M^h \gamma^h \alpha d + \beta^h \gamma^h d v^2) \tag{104}$$

In this model, the fuel consumption of vehicle of type h is determined by two variables: distance d and speed v, others are the coefficients specific to vehicle type h. The first part of E^h is linear in the running time, the second part is linear directly in the speed s, but the last part is quadratic in s.

Koç *et al.* [39] [40] apply this model to calculate the fuel consumption of a vehicle of certain weight from one departure node to the destination node, and use it as a parameter in the objective of the optimization model.

This fuel consumption model, comparing to the previous model from Jaumard *et al.* [31], does not ignore the segment of linearly on the traveling time, but lack another segment linearly on the speed. So their fuel consumption curve is not u-shape.

We plan to use the full model of equation (103) with all the four components. The value of those coefficients can be found in AREMA manual [11] and Radford [47]:

We will use the parameter values suggested in Condie [21], which comes from [11]:

- $A = 1.5 \cdot W + 18 \cdot N$ where
  - N = number of axles
  - W = weight of the train
- $B = 0.05 \cdot W$

Table 5: Case Study's Parameters

Type	Parameter	Value
Car	Weight	30/90/131.5 *tons*
Car	Canadian Streamlining coeff.	5.0
Loco.	Weight	212 *tons*
Loco.	Canadian Streamlining coeff.	leading 24.0, trailing 5.5
Loco.	Cross-sectional Area (a)	160 ft^2
Train	Number of Cars	75
Train	Number of Locomotives	3
Train	Number of axles N	6 for each locomotive

- $C = \dfrac{C_a \cdot a'}{10,000}$ where
  - C_a = Canadian Streamlining coefficient. It differs depending on whether the equipment is leading or trailing.
  - a' = the cross-sectional area of locomotive or car in square feet

Figure 16 shows the fuel consumption rate calculated by this model, under different speed for trains with 3 locomotives and 75 cars which has 3 different weights.

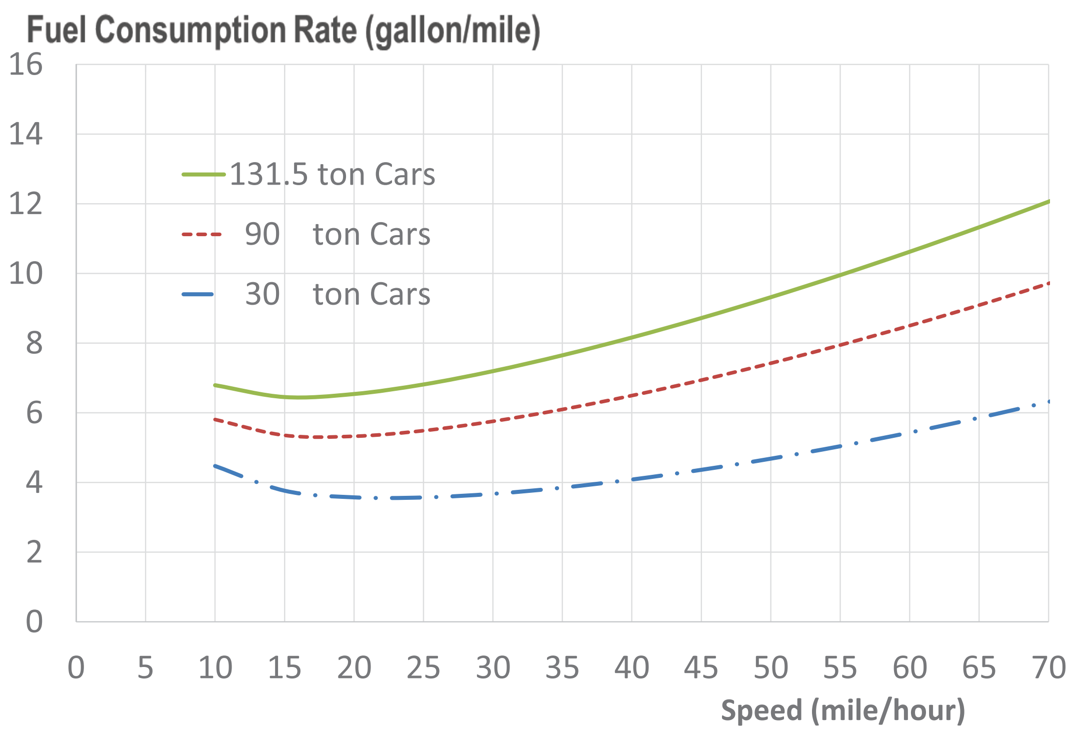


Figure 16: Fuel Consumption Rate Vs. Speed for Train with 3 Locomotives and 75 Cars

Based on the model, we do a calculation using these three types of train, which have the total weight of 2,886-ton, 7,386-ton, and 10,498.5-ton, have the fuel consumption from Montreal to Toronto (335 miles) showed in Figure 17.

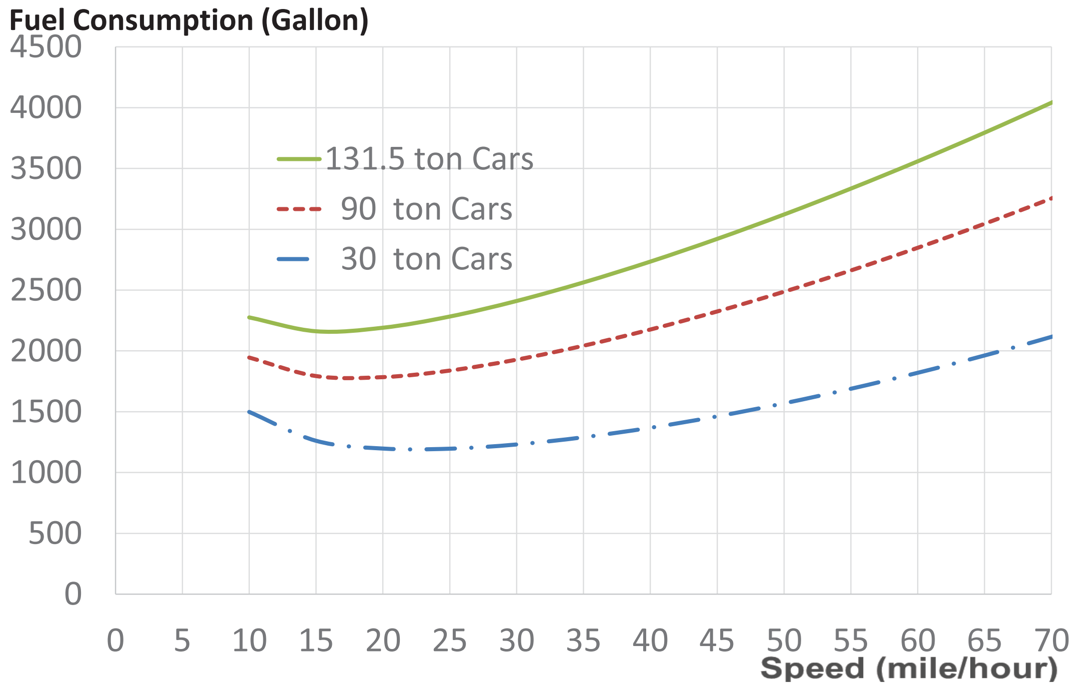


Figure 17: Total Fuel Consumption in Montreal-Toronto (335 miles)

In West Canada, the freight train usually has to travel through mountain area, e.g., the Rockies. The model needs to consider the energy for a train to climb over 3,500 feet. So we apply the gravity force and height factors to the model in Equation (103). The new model, i.e., Equation (105), considers the potential energy which is calculated by WH (W: weight in lbs, H: height in mile) and converts to the fuel consumption in gallon.

$$E(gal) = a \cdot \frac{d}{V} + b(A + BV + CV^2)\frac{d}{308} + \frac{WH}{308} \tag{105}$$

An official document from Canadian National Railway Company (CN) [4] provides this Figure 18, and indicates that because of the higher altitude, the fuel consumption from Seattle to Chicago is 104% more than Vancouver to Chicago with similar distance. Using our fuel consumption model, we can get the same result from the parameters given by CN with our 10,498.5-ton at speed of 30mph.

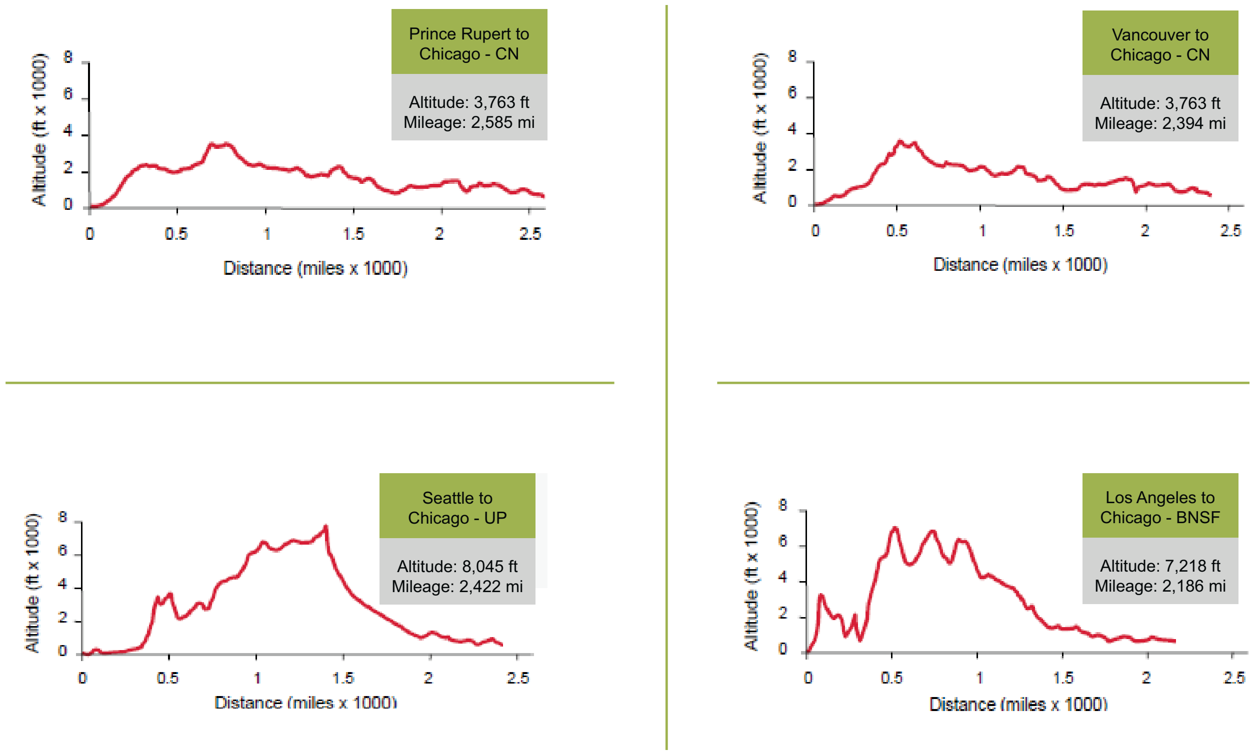


Figure 18: Railway Altitude Variance

4.5.2 LAP with Fuel Consumption Model

For our current model, the fuel consumption model works in the pricing problem, and for RMP, each consist travel plan $s \in S$ will have the attribute FuelCost_s for its fuel consumption:

So the objective of RMP now is

$$\sum_{s \in S} \text{FuelCost}_s \, z_s \tag{106}$$

The RMP objective has an issue that it tend in breakup the consist travel plans in to single trains, and so bring back the consist busting again, if we do not make any limits on the model. Based on the model and results on stage 1, we add Constraint (107) to set an upper bound to the number of consist travel plans allowed in the solution so to keep the reasonable number of consist busting.

$$\sum_{s \in S} z_s \leq UB. \tag{107}$$

For the pricing problem, the fuel cost for each consist travel plan it generated, $\text{FuelCost}_s = \sum_{\ell \in L^T \cup s} \sum_{b \in \beta} c_\ell^b \cdot x_\ell^{b,s}$.

Here we introduce new variables and parameters:

$b \in \beta = \{1, 2, 3, ..., 14\}$ represents the 14 speed choices from 10 to 70 mph that train selects for traveling. For each level b, the value of speed is σ^b.

$x_{\ell,p}^b = 1$ represents for consist travel plan path p, the average speed σ^b is used for the selected train link ℓ, 0 otherwise.

c_ℓ^b provides the value of fuel consumption for train link $\ell \in L^T$ under the speed level b. It is calculated via the fuel consumption model in equation (103) or (105) using the train and track parameters. Note it is not a decision variable set, but a set of parameters.

From Figure 16 we know the fuel consumption rate curves are convex, so when this curve is applied directly in a MILP optimization model, the objective will be nonlinear for decision variable of speed v. To linearize it, we apply the method proposed by Bektaş and Laporte [15] that discretizes the speed variables. We break the continuous speed into several speed levels $\beta = \{1, 2, ..., b, ...\}$, where in each level b, a value of average speed σ^b is set instead of the speed range of current interval, as well as the fuel consumption value c_ℓ^b for train ℓ.

And so now the reduced cost, i.e., objective of pricing problem is:

$$\overline{\text{COST}}_2 = \text{SIZE}(P) + \sum_{p \in P} \sum_{\ell \in L^T} \sum_{b \in \beta} c_\ell^b \cdot x_{\ell,p}^b - \overline{\text{COST}}_0. \tag{108}$$

And also we add these two additional constraints for the traveling speed and fuel consumption:

$$(d_\ell / \sigma^b) \cdot x_{\ell,p}^b \leq \sum_{v \in V_\ell^{\text{DST}}} Time(v) \cdot \alpha_{\ell,v,p}^{\text{DST}} - \sum_{v \in V_\ell^{\text{SRC}}} Time(v) \cdot \alpha_{\ell,v,p}^{\text{SRC}} \qquad p \in P, \ell \in L^T, b \in \beta \tag{109}$$

$$\sum_{b \in \beta} x_{\ell,p}^b = \sum_{v \in V_\ell^{\text{DST}}} \alpha_{\ell,v,p}^{\text{DST}} \qquad p \in P, \ell \in L^T \tag{110}$$

Constraints (109) and (110) work combined for the liberalization of speed that guarantees the running time of a selected train, determined by the average speed of certain speed level, should be no more than the time length between rescheduled source and destination nodes. And if train is not selected, the running time and speed should both be zero, otherwise they both should have a positive value.

The whole model can be solved by a general-purpose solver, e.g., Ilog Cplex.

4.6 Numerical Results

Now we have two models for LAP, the first one, original model(marked as LAP-Orig) from Jaumard *et al.* [33], focus on the minimization of locomotive fleet size andor the number of consist busting by given train schedule. The second model, LAP-Fuel, allows to relax the given trains' departure/arrival time as well as to integrate with the fuel consumption model, in order to optimize the fuel consumption of the locomotive fleet.

4.6.1 Data Instances

We use the entire CPR railway network, and the train schedule from CPR real operation. The test data set comes from different time section from given schedule, from 1-day to 2-week period. We set the maximum time length to two weeks because that the normal train from coast to coast takes around 1 week and the round trip is 2 week. For the maximum 2-week test scenario, there are 1,750 trains. The locomotives fleet contains 9 different types.

Now we have two models for LAP, the first one, original model(marked as *LAP-Orig*) from Jaumard *et al.* [33], focus on the minimization of locomotive fleet size andor the number of consist busting by given train schedule. The second and current model, *LAP-Fuel*, allows to relax the given trains' departure/arrival time as well as to integrate with the fuel consumption model, in order to optimize the fuel consumption of the locomotive fleet. We run these two models with the same six data set.

We use CPlex 12.6.1 as the ILP solver, which runs in a computer with 40-cores, 1TB memory. For the ϵ-optimal solution of ILP from optimal LP solution, we apply a simple rounding off procedure.

4.6.2 Computational Comparison of the Different Models/Algorithms

Table 6 lists the results of our current model: *LAP-Fuel*. We provides the commotional time, number of columns generates by PP, number of columns finally selected (initial and new generated), total locomotive fleet size required, fuel comsumptionsaving, and the gap gap of ϵ-optimal solution of ILP. In addition, we compare the solutions for these two different models in Table 7, for two scenarios: 1-week and 2-weeks. Our current fuel saving model is able to save more than 10% of fuel, with extra locomotive requirement.

# Trains	Time Length	Comp. Time	# Columns		GAP (%)	Loc. Req.	Fuel	
			Generated	Selected			Consump.	Sav.(%)
113	1 day	13m23s	531	93	1.99	256	535,351	3.56
357	3 days	30m59s	733	306	2.78	590	1,617,372	5.13
593	5 days	3h57m43s	1,276	507	2.00	943	2,550,183	7.71
862	7 days	13h45m19s	1,753	716	1.21	1,175	3,462,703	9.92
1,230	10 days	28h44m69s	2,609	1,086	0.32	1,327	5,386,609	10.35
1,750	14 days	32h51m16s	3,522	1,508	0.30	1,484	8,483,472	11.62

Table 6: Computational Results for Different Scenarios

# Trains	LAP Model	Comp. Time	# Columns		# Rounds	Loc. Req.	Fuel Consump.	Fuel Saving(%)	GAP (%)
			Generated	Selected					
7-day 862-train	Orig.	11h03m21s	1,543	506	4	962	3,844,030	N/A	1.24
	Fuel	13h45m19s	1,753	716	5	1,175	3,462,703	9.92	1.21
14-day 1,750-train	Orig.	26h29m31s	1,835	1,294	4	1,289	9,598,060	N/A	1.82
	Fuel	32h51m16s	3,522	1,508	5	1,484	8,483,472	11.62	0.30

Table 7: Computational Comparison of the Different Models

# Trains	Time Length	Avg. Travel Plan Size	Avg. Loc. Active Time	Loc. In-Consist Average Idle Time	Fuel Sav.(%)
113	1 day	1.22	12h03m	7h41m	3.56
357	3 days	1.16	22h25m	7h45m	5.13
593	5 days	1.17	2d07h19m	12h24m	7.71
862	7 days	1.20	3d17h17m	1d07h45m	9.92
1,230	10 days	1.13	5d05h12m	2d00h26m	10.35
1,750	14 days	1.16	7d12h03m	2d12h19m	11.62

Table 8: Characteristics of LAP Solutions for Different Scenarios

# Trains	LAP Model	Consist Travel Plan Size		Avg. Loc. Active Time	Loc. In-Consist Idle Time		Fuel Sav.(%)
		Average	Maximum		Average	Maximum	
7-day 862-train	Orig.	1.65	10	3d05h10m	9h38m	3d18h00m	N/A
	Fuel	1.20	11	3d17h17m	7h45m	2d04h41m	9.92
14-day 1,750-train	Orig.	1.35	13	6d10h03m	18h53m	8d03h19m	N/A
	Fuel	1.16	12	7d12h03m	16h19m	6d22h51m	11.62

Table 9: Characteristics of Solutions of Different Models

4.6.3 Characteristics of LAP Solutions

Table 8 lists several aspects of results from industry view, i.e., the average train number per consist travel plan, average locomotive running time, and average locomotive idling time within consist travel plan. Also we compare the solutions quality for these two different models in Table 9. It shows that to save fuel consumption, we pay not only extra locomotives, but also more running time per locomotive. For both groups, our fuel consumption model LAP-Fuel has shorter consist travel plan average size, in 14-day group, the maximum size of consist travel plan is also shorter for LAP-Fuel. It is because that model LAP-Fuel focuses on energy saving more than consist busting reduction, its solution prefer to assign exact power for train requirement, which leads to shorter consist travel plan, whereas model LAP-orig uses longer consist travel plan to optimize consist busting.

4.6.4 Analysis of Solution

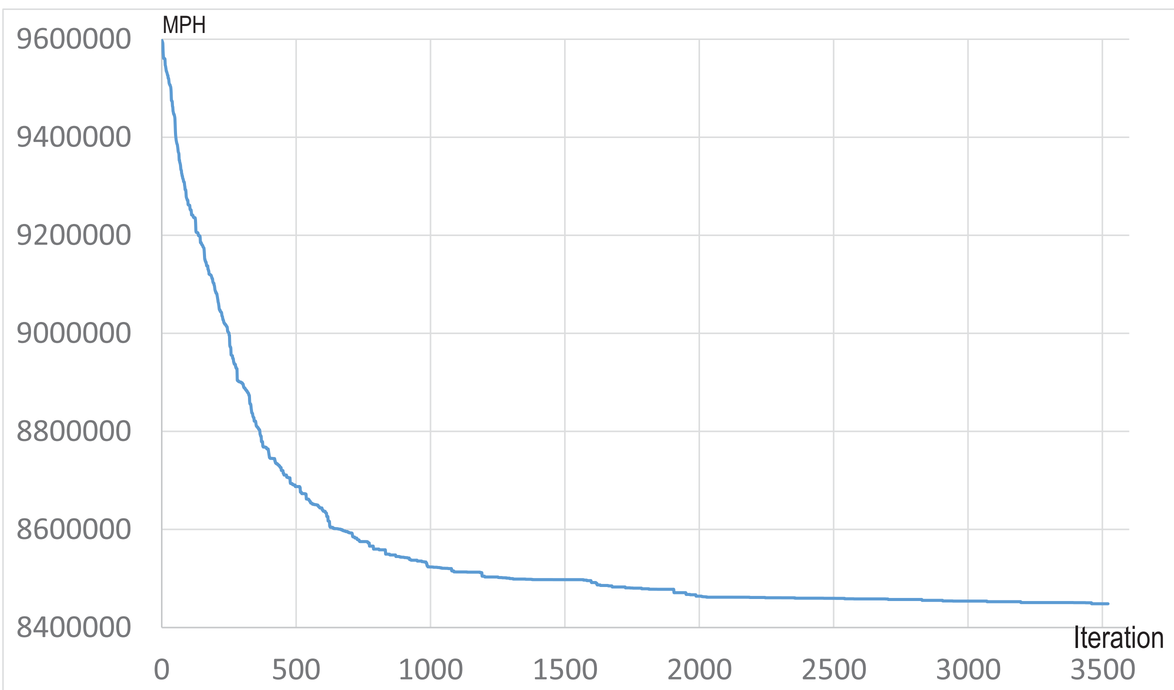


Figure 19: Fuel Consumption Convergence: 1,750-Train Scenario

Figure 19 shows the convergence rate curve during the problem solving. In Figure 20 two figures are merged together with the same x-axis which is the train sequence number. The upper figure compares original and rescheduled speed, the lower figure shows the fuel consumption comparison before and after rescheduling.

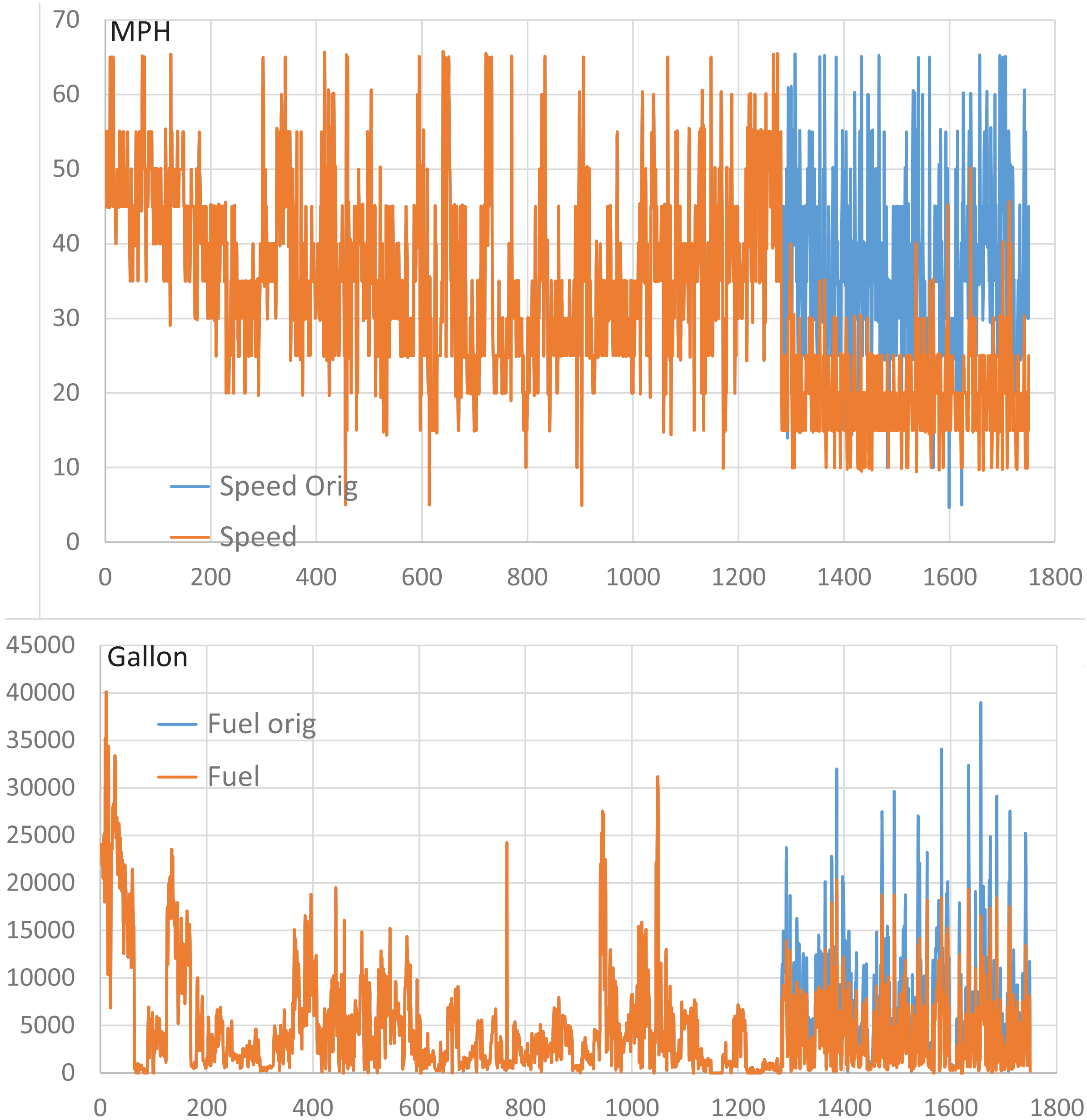


Figure 20: Train Rescheduling: 1,750-Train Scenario

In the figure, the first part of trains have unchanged speed, and in the second part most of the trains have speed adjustment. The reason is that, to draw the figure, we go through to check the selected column set of the final solution by RMP. By default sequence, that set contains first the original columns then the new added columns from PP. The original columns only contains trains unchanged, and the columns from PP have all the rescheduled trains with some unchanged trains. That's why the rescheduled trains only located on the second part of the figure. The fuel consumption comparison part has the same situation.

4.7 Conclusion

We propose an optimization model which integrates fuel consumption model to the locomotive assignment problem (LAP) model, and allows train scheduling, to reduce fuel consumption in freight train transportation area. With column generation decomposition, our new model in reasonable time can solve the given scenario up to 2-week, 1,750 train, based on the real world freight train operation of Canadian Pacific Railway, and achieves 11% of fuel consumption saving.

There might be a need for revised fuel/energy model taking into account the latest locomotive with new fuel saving equipments in the market. For cars, now the industry also optimize the aero dynamic characteristics. It is very challenging to build a very accurate predictive model matching the new and various parameters of locomotives and cars, but it demands the huge data support from cooperation with industry, and tie validation is costly. We hop our next project will do some work in this area.

Chapter 5

Conclusions and Future Work

In this thesis, we study two areas of train management, the locomotive assignment problem (LAP) and locomotive/train consumption problem. The results of the thesis have been published or submitted in [34, 37, 33, 52, 53].

5.1 Thesis Conclusions

Freight train transportation becomes more and more important by its high energy efficiency and safety advantage. But it also suffers from the high investment and operation cost, especially for locomotive fleet and the energy consumption of the operation. To optimize the number of locomotives needs and its operation cost, many research works have been done for locomotive assignment problem (LAP), including exact optimization model, heuristics, etc. However, most of the past exact LAP models can not cover all the rules needed in railway locomotive operation, e.g., locomotive maintenance and/or consist busting avoidance, and they also have scalability or solution accuracy issues. There are also some models that allow to adjust the train schedule within a given time window so to make the locomotive assignment more flexible, but none of them provide how to define the reasonable time length of time window and whether the rescheduled train will affect other trains to keep the train schedule feasible. Finally, there are some train fuel consumption model, but no past research tried to combine this kind of model with LAP model so to optimize the total train/locomotive energy consumption by locomotive assignment and train rescheduling.

In this thesis, we study the LAP and related problems, and give answer to the main research question:

How to optimize the locomotive assignment for a given train schedule?

We decompose the main question into four sub-questions. The first question focuses on the consist busting avoidance.

1. How to evaluate the consist busting occurrence in LAP ILP model so to optimize it?

In Chapter 2, we propose a consist travel plan based optimization LAP model, which use the number of total consist travel plans to represent the the consist busting occurrence. The reason is that when a new consist travel plan is applied, it means that a consist busting has been executed.

The following question is for the locomotive maintenance constraints.

2. How to allow maintenance process in LAP ILP model?

The propose a consist travel plan based optimization LAP model in Chapter 2 has a special flow conservation constraints (3) & (4) that allow locomotive change the status from critical to normal, so we can reuse the maintained locomotive after the maintenance/shop flow in our model.

With column generation decomposition technique, our LAP model solve the maximum of data instances over 2-week's train schedule in affordable time. The solution makes compromise between the locomotive fleet size and consist busting avoidance, as well as sends 20% of locomotives which are critical to maintenance shops under maintenance rule. It reduces the number of consist travel plans from 1,750 to 1,294 which means that in worst case it saves 26% of consist busting occurrences.

The third question is about the time efficiency of LAP model.

3. Can the computational time of LAP model be accelerated?

To further reduce the computational time, we study various strategies of acceleration for column generation, and draw the inspiration from one of them that can generate several columns with negative reduced cost. In Chapter 3 we establish a new column generation architecture with the generation of multiple columns per pricing problem call, each of which is the optimal solution in a pre-processed, reduced sub-network from the general time-space train scheduling network. Keeping almost the same quality of final solution, our new multi-CG LAP model reduces the computational time at least 60% and the maximum time saving is 93%. Our new CG architecture could be applied to the general network flow problem which has potential network decomposition into several overlapped sub-networks and each for a pricing problem.

Finally, in response to the tendency of energy saving and the needs of railway industry, we study the fuel consumption problem as the fourth question.

4. How to optimize train/locomotive energy usage in train schedule and locomotive assignment level?

In Chapter 4, as the pre-requirement of LAP with locomotive fuel consumption model, we allow train rescheduling in a time window. In order to avoid excessive change of the given train schedule to keep it still feasible, i.e., the adjustment of one train does not change the meeting time/siding with other trains, we develop a conflict-free pre-process algorithm. During this stage, we consider the train meeting time/position, and the maximum/minimum speed limit, so to establish a set of reasonable departure/arrival time windows for each train. Then our LAP model with a linearized fuel consumption optimization can save up to around 12% of energy consumption comparing to the results from our previous LAP model in Chapter 2 & 3. The result comes from the semi-artificial initialization data we generates from train schedule of CPR. Suppose enough details of real train/locomotive parameters and operation data are available, our model can provides a more precise solution of energy consumption for analysis.

5.2 Future Work

Our optimization LAP model applies time-based maintenance constraints, which is one of the general constraint used in the freight train industry. It set a fixed shop time period (usually three months) for locomotive. On the other hand, there is a more flexible type of maintenance constraint, which is milage-based, that counts the travel distance of locomotive and only requires it to go to a shop when the total distance reaches the limit. We can apply the milage-based maintenance constraint to the LAP model, and analysis the benefit and efficiency. In addition, the distributed power of consist configuration will also be considered.

Our CG acceleration architecture now works fine in the LAP model. In addition it has the potential to be applied to the similar format of CG models. The first target is the network flow models with CG, which has the type of network that can be reformatted to several overlapped sub-network. We plan to expand the new architecture to those models and analysis the time efficiency.

For the model of fuel consumption optimization on LAP and train scheduling level, our algorithm for conflict-free train rescheduling time window now keep the given train meeting time and siding. The advanced version of conflict-free algorithm should works with more detail of train schedule and railway network, and has the relaxation to allow train meeting in different time and siding out of given schedule. With the new, more flexible and reasonable time windows, our model will get further energy saving.

For the train fuel consumption model part, current models was developed decades ago, and can not reflect accurately the new design and technologies applied in locomotive/train fuel consumption rate. Also the advanced train fuel consumption model should consider the acceleration/dacceleration, and flop of tracks. Finally, the model should be fully validated by the real word train operation before being applied to the optimization model. In addition, with the future requirement of new energy, e.g., natural gas or electricity power train, we still need to update the current energy consumption model with extra validation as well.

Bibliography

[1] Analysis of railway fulfillment of shipper demand and transit times. QGI Consulting. http://www.qgiconsulting.ca/pdf/FulfillmentofShipperDemandandTransitTimes.pdf, 2010. [last visit on December 03, 2017].

[2] Rail freight service review - final report (tp 15042). Transport Canada. http://publications.gc.ca/collections/collection_2011/tc/T42-3-2-2011-eng.pdf, 2011. [last visit on December 03, 2017].

[3] The CP book. Canadian Pacific Railway Limited. http://www.cpr.ca/en/invest-in-cp/investor-book/Documents/2012-cp-investor-book.pdf, 2012. [last visit on December 03, 2017].

[4] Fuel program. Canadian National Railway Company. https://www.cn.ca/-/media/Files/Customer%20Centre/Customer-Centre-Documents/fuel-surcharge-en.pdf, 2012. [last visit on December 03, 2017].

[5] Railway locomotive inspection and safety rules. Transport Canada. https://www.tc.gc.ca/media/documents/railsafety/Locomotive-Rules-July-3-2015.pdf, 2015. [last visit on December 03, 2017].

[6] Canadian pacific annual report. Canadian Pacific Railway Limited. http://s21.q4cdn.com/736796105/files/doc_financials/Annual-Report/2016/CP_Annual-Report_2016.PDF, 2016. [last visit on December 03, 2017].

[7] DOT releases 30-year freight projections. United States Department of Transportation. https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/bts13_16.pdf, 2016. [last visit on December 06, 2017].

[8] Transportation in Canada 2016. Transport Canada. https://www.tc.gc.ca/eng/policy/transportation-canada-2016.html, 2017. [last visit on December 03, 2017].

[9] R.K. Ahuja, J. Liu, J.B. Orlin, D. Sharma, and L.A. Shughart. Solving real-life locomotive-scheduling problems. *Transportation Science*, 39(4):503–517, 2005.

[10] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, 1993.

[11] American Railway Engineering and Maintenance-of-Way Association. *Manual for railway engineering*, 2013.

[12] F. An and M. Ross. Model of fuel economy with applications to driving cycles and traffic management. *Transportation Research Record*, page 105, 1993.

[13] C.B. Barber, D.P. Dobkin, and Hannu H. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22(4):469–483, 1996.

[14] M. Barth, T. Younglove, and G. Scora. Development of a heavy-duty diesel modal emissions and fuel consumption model. *California Partners for Advanced Transit and Highways (PATH)*, 2005.

[15] T. Bektaş and G. Laporte. The pollution-routing problem. *Transportation Research Part B: Methodological*, 45(8):1232–1250, 2011.

[16] V. Cacchiani, A. Caprara, and P. Toth. Solving a real-world train-unit assignment problem. *Mathematical programming*, 124(1-2):207–231, 2010.

[17] V. Cacchiani, A. Caprara, and P. Toth. Models and algorithms for the train unit assignment problem. In *International Symposium on Combinatorial Optimization*, pages 24–35. Springer, 2012.

[18] V. Cacchiani, A. Caprara, and P. Toth. A lagrangian heuristic for a train-unit assignment problem. *Discrete Applied Mathematics*, 161(12):1707–1718, 2013.

[19] S. Chen and Y. Shen. An improved column generation algorithm for crew scheduling problems. *Journal of Information and Computational Science*, 10(1):175–183, 2013.

[20] V. Chvátal. *Linear programming. A series of books in the mathematical sciences*. WH Freeman and Company, New York, USA, 1983.

[21] R.A. Condie. Aerodynamic improvements for auto-carrying railcars. Master's thesis, Brigham Young University, 2014.

[22] J.-F. Cordeau, F. Soumis, and J. Desrosiers. A benders decomposition approach for the locomotive and car assignment problem. *Transportation science*, 34(2):133–149, 2000.

[23] J.-F. Cordeau, F. Soumis, and J. Desrosiers. Simultaneous assignment of locomotives and cars to passenger trains. *Operations research*, 49(4):531–548, 2001.

[24] G. Desaulniers, J. Desrosiers, and M.M. Solomon. Accelerating strategies in column generation methods for vehicle routing and crew scheduling problems. In *Essays and Surveys in Metaheuristics*, pages 309–324. Springer US, Boston, MA, 2002.

[25] D. Emerson. Pathways: Connecting Canada's transportation system to the world, volume 1. *Canada Transportation Act Review*, 2015.

[26] A. Fügenschuh, H. Homfeld, A. Huck, and A. Martin. Locomotive and Wagon Scheduling in Freight Transport. In *6th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS)*, volume 5 of *OpenAccess Series in Informatics (OASIcs)*, Dagstuhl, Germany, 2006.

[27] K. Ghoseiri and S.F. Ghannadpour. A hybrid genetic algorithm for multi-depot homogenous locomotive assignment with time windows. *Applied Soft Computing*, 10(1):53–65, 2010.

[28] Keivan Ghoseiri, Ferenc Szidarovszky, and Mohammad Jawad Asgharpour. A multi-objective train scheduling model and solution. *Transportation research part B: Methodological*, 38(10):927–952, 2004.

[29] T. Godwin, R. Gopalan, and T. Narendran. Tactical locomotive fleet sizing for freight train operations. *Transportation Research Part E: Logistics and Transportation Review*, 44(3):440–454, 2008.

[30] J.-L. Goffin and J.-P. Vial. Multiple cuts in the analytic center cutting plane method. *SIAM Journal on Optimization*, 11(1):266–288, 2000.

[31] B. Jaumard and T. H. Le. Optimizing freight train scheduling with respect to energy consumption. In *The Third International Conference on Railway Technology: Research, Development and Maintenance*, Stirlingshire, United Kingdom, 2016.

[32] B. Jaumard, T. H. Le, H. Tian, A. Akgunduz, and P. Finnie. An enhanced optimization model for scheduling freight trains. In *Joint Rail Conference*. American Society of Mechanical Engineers, 2013.

[33] B. Jaumard and H. Tian. Multi-column generation model for the locomotive assignment problem. In *16th Workshop on Algorithmic Approaches for Transportation Modelling,*

Optimization, and Systems (ATMOS), volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 6:1–6:13, Dagstuhl, Germany, 2016.

[34] B. Jaumard, H. Tian, and P. Finnie. Locomotive assignment under consist busting and maintenance constraints. In *Joint Rail Conference*, 2014.

[35] B. Jaumard, H. Tian, and P. Finnie. Locomotive assignment under consist busting and maintenance constraints. *Submitted*, 2014. `https://www.gerad.ca/en/papers/G-2014-54`.

[36] B. Jaumard, H. Tian, and P. Finnie. Best compromise in deadheading and locomotive fleet size in locomotive assignment. In *Joint Rail Conference*, 2015.

[37] B. Jaumard, H. Tian, and P. Finnie. A new model for locomotive assignment under consist busting. *submitted for publication*, 2017.

[38] S. Kasalica, D. Mandić, and V. Vukadinović. Locomotive assignment optimization including train delays. *PROMET-Traffic&Transportation*, 25(5):421–429, 2013.

[39] Ç Koç, T. Bektaş, O. Jabali, and G. Laporte. The fleet size and mix pollution-routing problem. *Transportation Research Part B: Methodological*, 70:239–254, 2014.

[40] Ç. Koç, T. Bektaş, O. Jabali, and G. Laporte. The impact of location, fleet composition and routing on emissions in urban freight distribution. *Transportation Research Part B*, 84:81–102, 2016.

[41] D. Maposa and S.D. Swene. Locomotive scheduling in freight transport at Mpopoma train station in Bulawayo for the southern region, Zimbabwe. *International Journal of Economics and Management Sciences*, 1(12):104–116, 2012.

[42] A. Mingozzi, M.A. Boschetti, S. Ricciardelli, and L. Bianco. A set partitioning approach to the crew scheduling problem. *Operations Research*, 47(6):873–888, 1999.

[43] S. Noori and S. F. Ghannadpour. Locomotive assignment problem with train precedence using genetic algorithm. *Journal of Industrial Engineering International*, 8(1):1–13, 2012.

[44] F. Piu. A mixed integer programming approach to the locomotive assignment problem. In *Swiss Transport Research Conference-(STRC)*, pages 1–42, 2011.

[45] F. Piu and M.G. Speranza. The locomotive assignment problem: a survey on optimization models. *International Transactions in Operational Research*, 21(3):327–352, 2013.

[46] W. Powell, B. Bouzaiene-Ayari, C. Cheng, R. Fiorillo, S. Das, and C. Lawrence. Strategic, tactical and real - time planning of locomotives at Norfolk Southern using approximate dynamic programming. In *Joint Rail Conference*, pages 491–500, 2012.

[47] R.W. Radford. Fuel consumption of freight trains hauled by diesel electric locomotives. *Journal of Engineering for Industry*, 105(2):75–87, 1983.

[48] S. Rouillon, G. Desaulniers, and F. Soumis. An extended branch-and-bound method for locomotive assignment. *Transportation Research. Part B, Methodological*, 40(5):404–423, 2006.

[49] M. Saddoune, G. Desaulniers, I. Elhallaoui, and F. Soumis. Integrated airline crew pairing and crew assignment by dynamic constraint aggregation. *Transportation Science*, 46(1):39–55, 2012.

[50] R. Sadykov, F. Vanderbeck, A. Pessoa, and E. Uchoa. Column generation based heuristic for the generalized assignment problem. *XLVII Simpósio Brasileiro de Pesquisa Operacional, Porto de Galinhas, Brazil*, 2015.

[51] C. Surapholchai, G. Reinelt, and H.G. Bock. Solving city bus scheduling problems in bangkok by eligen-algorithm. In *Modeling, Simulation and Optimization of Complex Processes*, pages 557–564. Springer, 2008.

[52] H. Tian and B. Jaumard. An enhanced decomposition scheme for locomotive assignment problem. *submitted for publication*, 2017.

[53] H. Tian and B. Jaumard. Minimizing fuel consumption for freight trains. *submitted for publication*, 2017.

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

[55] 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(2):281–292, 1997.

[56] K. Ziarati, F. Soumis, J. Desrosiers, and M.M. Solomon. A branch-first, cut-second approach for locomotive assignment. *Management Science*, 45(8):1156–1168, 1999.