APPENDIX: An approach for the railway multi-territory dispatching problem

Conrado Borraz-Sánchez
Data&Analytics Center, KPMG-Lighthouse LLP, Knoxville, TN-37922
cborrazsanchez@kpmg.com

Diego Klabjan
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL-60208
d-klabjan@northwestern.edu

Alper Uygur
BNSF Railway, Fort Worth, TX-76131
alper.uygur@gmail.com

Abstract: In this research, we address the railway multi-territory dispatch planning (RMTDP) problem. The goal of the RMTDP problem is to find the optimal movement of trains across consecutive dispatch territories; and it is one of the major challenges that decision makers face on a daily basis. It ideally takes into account the correct placement of maintenance windows, remaining capacity of terminals and availability of train crews among other critical aspects such as locomotive balance, fueling locations and inspections. Although these train movement plans are made at the corridor level, which is comprised of several dispatch territories, when it comes to execution, the meet-pass decisions are made at the individual dispatch territories. This notion causes disruptions and misalignment at the boundaries of dispatch territories. The approach in this paper aims at finding a holistic conflict-free master plan by optimally matching train line-ups at territory boundaries and smoothly routing trains through bottlenecks.

We propose an efficient solution approach that iteratively constructs a master scheduling plan while minimizing the amount of train delays within a given planning horizon. This is accomplished by designing a time-space network model to identify feasible schedules and developing a mathematical programming-based heuristic to solve the underlying model. A thorough computational study shows the effectiveness of our heuristic approach, as we report reasonable average run times of 3.0 and 6.5 minutes to solve instances of moderate to large size problems, respectively.

The results obtained from the algorithm using test snapshots from a Class I Railroad company have been shown to assistant chief dispatchers and have received encouraging feedback for applicability.

Appendix

A. A Discrete-Time Modeling Approach For A Single Dispatching-Territory

A.1. Problem Description

Railway companies divide their large freight service area, called *corridors*, into multiple individual territories as an operational strategy to effectively plan the movements of trains over a given planning horizon. Each territory is a centralized traffic control area that is under the governance of a train dispatcher who operates it individually from any other adjacent territory. The boundaries of an individual dispatching territory are usually outlined by the start of sections known as *terminals*, which are complex sections controlled by terminal trainmasters or yardmasters. Note that the end points of a dispatching territory along with a terminal conform two out of the three main components used to create *the stitching network areas*, which are used in the present work to holistically optimize an entire corridor. (The third component is a virtual linkage designed to “stitch” the two different network configurations, dispatching territories and terminals, as described in Section **??**.)

A dispatching territory typically consists of 100 to 200 miles of single or double tracks with multiple intermediate *sidings*, also called *stations*. Sidings serve of two main purposes: (a) avoid collisions when trains traveling on the same track but in opposite directions *meet* and cross each other –i.e., one of the trains uses the mainline while the other uses or waits on a siding; and (b) increase the utilization of tracks (efficiency) by letting a fast or high priority train overtake another train traveling in the same direction, which alludes to a *pass* event –i.e., the latter train waits on a siding to permit the former train to pass.

The mathematical formulation introduced here follows [**?**] and [**?**] and adopts a discrete time modeling approach to optimally resolve all the expected events (the decision of movements of trains) within an individual dispatching territory in a specified time frame (8 to 12 hours), leaving out any potential event required beyond its boundaries. We refer to this optimization problem as *the single dispatching territory problem* (SDTP).

SDTP network configuration Without loss of generality, the network configuration of the SDTP is based on two main network components that consistently alternate along the territory: sidings or stations, and *segments*. Segments are defined as mainline tracks between two adjacent stations with no intermediate locations where a meet & pass event can occur, thus trains traveling in opposite directions are forbidden from using the same segment simultaneously. Trains traveling in the same direction, however, are allowed to trail each other within segments as long as a minimum required distance, called the *trailing headway*, is met.

Dispatchers usually specify the trailing headway in terms of *blocks*, which are portions of tracks between two adjacent signals on a segment.

The SDTP network also considers *control points*, which are located at any junction where a single track splits into multiple tracks or vice versa. Control points represent railroad switches that guide trains from one track to another.

Safety rules The SDTP requires that at most one train is observed in each block at any time. For safety reasons, this requirement is further extended by dispatchers by separating two successive trains by at least one unoccupied block so that the trailing train can move safely at its nominal speed with sufficient stopping distance.

To model the SDTP's separation rule, we treat every two adjacent blocks as a section and permit at most one train (traveling in either direction) to occupy a section at any time. More precisely, if a segment has K blocks, indexed as blocks $1, 2, \ldots, K$, we define blocks k and $k+1$ as section k for $k = 1, 2, \ldots, K-1$.

Another safety requirement concerns the minimum required time separation between trains crossing a control point. To ensure safe operation, successive trains passing through a control point (in either direction) must be separated by a minimum required time (typically, five to seven minutes). This requirement is referred to as the *control point headway*.

Assumptions The model formulation assumes specific properties of each train to be known parameters like, its travel direction, the sequence of stations or segments that it must traverse, its priority (used to decide the importance or weight for maximizing average velocity), any arrival time windows or hard time window, and the train's traversal (travel) time on each segment or station that it passes through.

Trains can only wait at stations and not on segments, and at most one train per track at any time. Based on a train's entry and traversal times, waiting time restrictions, track availability, and any explicit arrival time window requirements at intermediate stations or at the end of the territory, it is possible to calculate the time periods at which the train can enter stations or segments using control points.

We assume for simplicity that, at every station, the train's traversal time (including any track crossover time) is the same on all the tracks at that station. With this assumption (which largely holds in practice, particularly when we take into account the control point headway requirements), we do not need to distinguish between assignments and movements of trains along different tracks within the station.

The model largely focuses on single-track territories in which every segment has a single bi-directional track; with modest changes it also extends to multitrack segments. The notation used in the formulation of the SDTP is presented next.

A.2. Notation

The formulation of the SDTP considers a planning horizon that is split into fine-grained time intervals or *periods* (1 or 2 minutes). Let H denote the total number of time periods in the planning horizon, indexed as $t = 1, 2, \ldots, H$.

The STDP network configuration considers a set of stations (sidings) $s \in S$, a set of segments (mainline tracks) $m \in M$, and a set of control points $p \in P$ (switches). Let $e \in S \cup M$ (stations and segments) represent an edge in the territory. Each track segment m is partitioned into one or more *sections*, denoted as G_m, to ensure that trailing trains (traveling in the same direction) maintain sufficient inter-train headway or distance. Each section $s \in G_m$ can be occupied by no more than one train at any time. The travel direction for each train $q \in Q$ is denoted as $+$ (eastbound or northbound) or $-$ (westbound or southbound.)

Let o^q and d^q respectively denote train q's starting and ending stations in the territory. Assume, without loss of generality, that these starting and ending locations are stations where train q can wait (e.g., the terminals for future trains that pass through the territory, or intermediate stations for trains that are already in the territory). We define θ_p as the minimum required control point headway (in time periods) between two successive trains passing through control point p in either direction.

For every edge e and train $q \in Q_e$, we are given the time τ_e^q (in number of periods) for train q to traverse that edge. So, if α_{st} denotes the number of available parallel tracks at station s at time t, we can simply impose an aggregate capacity of α_{st} on the number of trains that are within the station in period t. We can readily extend the model to permit varying traversal times on different tracks, but at the expense of adding more decision variables.

Let $T_e^q(T_p^q)$ be the subset of periods in which train $q \in Q_m$ can enter edge e (control point p). At the origin station $s = o_q$, for trains that are already within the territory at the start of the planning horizon, this time window may include only the time period at which the train entered that station (with appropriate adjustments to traversal time so that entry time is at or after time zero).

On the other hand, for future trains that will enter the territory later, the time window may include all periods until the end of the horizon (including a dummy period $H+1$) if the train is permitted to wait at its starting location. Since trains cannot wait on segments, for every segment m and each section $g \in G_m$, we can determine the time needed for a train to enter (and leave) section g after it enters segment m. Specifically, let $\sigma_g^q(\lambda_g^q)$ be the time needed for train q to travel from the beginning of segment m to the beginning (end) of section $g \in G_m$.

For the sake of readability, all the notation (indices, sets, and parameters) used in the model formulation is listed in Tables 1–2.

A.3. Variables and Objective

Variables The SDTP's main dispatching decisions are when each train should enter each segment or station, and whether it should wait at a station. To capture these decisions, we define the

Table 1 – MIP model notation: Sets and indices

Symbol	Description
S	set of stations in the territory
M	set of segments in the territory
Q	set of trains to be dispatched
P	set of control points in the territory
s	index of a station
m	index of a segment
e	index of an edge
t	index of time
g	index of a section
p	index of a control point
$+,-$	indices of train directions

Table 2 – MIP model notation: Parameters

Symbol	Description
H	number of time periods in the planning horizon
G_m	set of sections in segment m
θ_p	control point headway (number of periods) at control point p
α_{st}	number of available tracks in station s at time t
o^q	origin edge of train q
d^q	destination edge of train q
τ_e^q	traversal time of train q on edge e
Q_e^k	set of trains that will enter edge e in direction k, where $k \in \{+,-\}$
Q_e	set of trains that will enter edge e
Q_p	set of trains that will use control point p
f_{st}^q	cost for train q to wait at station s at time t
c_{et}^q	cost for train q to enter edge e at time t
T_e^q	set of time periods that train q can enter edge e
T_p^q	set of time periods that train q can enter control point p
σ_g^q	time (number of periods) for train q to travel from the beginning of segment m, where $g \in G_m$, to the beginning of section g
λ_g^q	time (number of periods) for train q to travel from the beginning of segment m, where $g \in G_m$, to the end of section g
b_e^q	edge that train q travels on before it enters edge e
b_p^q	edge that train q travels on before it uses control point p

Table 3 – MIP model notation: Variables

Symbol	Description
x_{et}^q	$= 1$ if train q enters edge e at time t; 0 otherwise
y_{st}^q	$= 1$ if train q waits at station s (after entering station s) at time t; 0 otherwise.

following decision variables (see Table 3): $x_{et}^q = 1$ if train q enters edge e at time t, and 0 otherwise; $y_{st}^q = 1$ if train q waits at station s (after entering station s) at time t, and 0 otherwise. Note that we define variables based on segments, rather than on blocks, as suggested in [?] and [?], thereby dramatically reducing the number of variables.

Objective The goal of the STDP model is to optimize track utilization within the entire territory. This objective can be stated as (A) a minimization problem of the weighted waiting time of trains, or (B) a maximization problem of the weighted velocity of trains.

Let us apply objective (A) above so that we minimize the weighted waiting time. We will show later how to set the objective function coefficients so that objective (B) can be applied, and thus, the weighted velocity is maximized instead.

Assume ε^q is the weight for train q such that we can set $f^q_{st} = \varepsilon^q$ for all $s \in S$ and $t \in T^q_s$, $c^q_{et} = (t - t^q)\varepsilon^q$ if $e = o^q$, and $c^q_{et} = 0$ otherwise. Let τ^q represent the total runtime for train q to traverse all the stations and segments in the territory, and D^q represent the total length of all stations and segments combined.

Let t^q_e be the earliest expected time for train q to enter segment or station e, which can be achieved when train q does not wait before entering a. If train q enters d^q at time t, the total time it spends in the territory is $\tau^q + t - t^q_{d^q}$, and its velocity is $D^q/(\tau^q + t - t^q_{d^q})$. Accordingly, the objective function coefficients can be set as follows:

$$\overline{f}^q_{st} = 0; \overline{c}^q_{et} = -\varepsilon^q D^q/(\tau^q + t - t^q_{o^q}) \text{ if } e = d^q, \text{ and } \overline{c}^q_{et} = 0 \text{ if } e \neq d^q. \tag{1}$$

Coefficients (1) use $c^t_{d^q}$ to record the velocity of train q when q arrives at destination d^q. Alternatively, we can set the coefficients in the following way in order to maximize trains' weighted velocity:

$$\begin{aligned} &\hat{f}^q_{st} = \varepsilon^q D^q/(\tau^q + t - t^q_{o^q} + 1) - \varepsilon^q D^q/(\tau^q + t - t^q_{o^q}), \\ &\hat{c}^q_{et} = -\varepsilon^q D^q/(\tau^q + t - t^q_a) \text{ if } e = o^q, \\ &\hat{c}^q_{et} = 0 \text{ if } e \neq o^q. \end{aligned} \tag{2}$$

Theorem 1. *Objective function coefficients (1) and (2) are equivalent.*

Setting objective function coefficients according to (1) requires each train to enter its destination, which may not be achieved within the planning horizon. On the other hand, it is not necessary to send the train to its destination to get a valid velocity under coefficients (2). If a train ends up in certain location before its destination, it is assumed to run unimpeded from that location to its destination.

Computational experiments show that commercial MIP solvers like CPLEX can solve the model with coefficients (2) faster than the model with coefficients (1). Thus, coefficients (2) are used in subsequent discussions. Note also that to model the implicit time window requirement, we define $c^q_{et} = \hat{c}^q_{et} + \xi e t^q$ and use ξ^q_{et} to reflect the preference for train q to enter edge e at time t.

A.4. Mathematical Formulation

Given the notation defined in the previous section, the SDTP model can be formulated as follows:

$$z = \min \sum_{e \in S \cup M} \sum_{q \in Q_e} \sum_{t \in T^q_e} c^q_{et} x^q_{et} + \sum_{s \in S} \sum_{q \in Q_s} \sum_{t \in T^q_s} f^q_{st} y^q_{st} \tag{3}$$

subject to

$$\sum_{t \in T^q_e} x^q_{et} = 1, \qquad \forall q \in Q, e = o^q, \tag{4}$$

$$x^q_{mt_1} = x^q_{st_2}, \qquad \forall s \in S \backslash \{o^q\}, q \in Q_s, t_2 \in T^q_s, m = b^q_s, t_1 = t_2 - \tau^q_m, \tag{5}$$

$$x^q_{st_1} + y^q_{s,t_2-1} = x^q_{mt_2} + y^q_{st_2}, \qquad \forall m \in M \backslash \{o^q\}, q \in Q_m, t_2 \in T^q_s, s = b^q_m, t_1 = t_2 - \tau^q_s, \tag{6}$$

$$\sum_{q \in Q_s} y^q_{st} + \sum_{q \in Q_s} \sum_{t_1 = t - \tau^q_s + 1}^{t} x^q_{st_1} \leq \alpha_{st}, \qquad \forall s \in S, t = 1, \ldots, H, \tag{7}$$

$$\sum_{q \in Q_p : e = b^q_p} \sum_{t_1 = t - \tau^q_e + 1}^{t - \tau^q_e + \theta_p} x^q_{et_1} \leq 1, \qquad \forall p \in P, t \in T^q_p, \tag{8}$$

$$\sum_{q \in Q_m} \sum_{t' = t - \lambda^q_g + 1}^{t - \sigma^q_g} x^q_{mt'} \leq 1, \qquad \forall m \in M, g \in G_m, t \in T^q_m, \tag{9}$$

$$\sum_{\substack{q \in Q^+_m, \\ t' = t - \sigma^p_g}} x^q_{mt'} + \sum_{\substack{q \in Q^-_m, \\ t' = t - \lambda^p_g}} x^q_{mt'} \leq 1, \qquad \forall m \in M, g \in G_m, t \in T^q_m, \tag{10}$$

$$x^q_{et} \in \{0,1\}, \qquad \forall (q,e,t) \in A, \tag{11}$$

$$y^q_{st} \in \{0,1\}, \qquad \forall (q,s,t,t+1) \in A. \tag{12}$$

Eq. (3) represents the objective function that minimizes the trains' enter time at each station and segment plus the total waiting time expected of each train at the stations. As discussed earlier, the objective can be interpreted as maximizing the trains' weighted velocity, or minimizing the trains' total waiting time.

Constraint (4) enforces that each train enters the territory and is dispatched. Constraints (5) and (6) are typical flow conservation constraints found in any network transportation problem. Constraint (5) states that if train q enters station s at time t_2, it must have entered segment $m = b_s^q$ at time $t_2 - \tau_m^q$. Constraint (6) ensures that if train q enters the segment m or waits at its previous station s at time t_2, it must have entered station s at time $t_2 - \tau_s^q$ or waited at station s at time $t_2 - 1$.

Constraint (7) captures the station capacity requirements to ensure that the number of trains moving and waiting in a station s at any time t should not exceed the number of tracks available at the station. Since we assume that the travel time through a station is the same for all tracks at the station, we do not define separate variables for the movement and waiting of trains on each parallel track within a station; using a single variable for a train's movement or waiting at a station not only reduces the problem size but also avoids symmetry in the feasible solution space. A post-processing procedure may be considered in order to assign trains to specific tracks within a station. Constraint (8) enforces the control point headway requirement: at most one train can pass through control point p for every θ_p time units.

Constraint (9) ensures that each section is occupied by no more than one train in any time period so as to prevent any deadlock situation. Hence, trains traveling in opposite directions are forbidden from appearing in a particular section at the same time. We refer to this property as the unidirectional movement property.

Since a section consists of two adjacent blocks, constraint (9) also guarantees that two trailing trains are separated by at least one unoccupied block. This constraint can also ensure that no meeting or overtaking occurs inside segments. Observe that, if a train overtakes or meets another train inside a segment, they have to appear in some section of that segment at the same time which would violate the unidirectional movement property; therefore, ensuring that no more than one train uses each track section at any time can prevent meet & pass events inside segments. Finally, constraint (10) prevents trains from crossing each other at the boundary point of two adjacent blocks. As shown in [**?**], the formulation are not valid without constraint (10).

A.5. Computational results

Preliminary computational experiments using a commercial solver, Cplex 12.4, showed modest results when solving the single dispatching territory problem (SDTP) with the MIP model presented in the previous section. More precisely, for SDTP-scenarios of more than 200 miles of tracks with a number of trains ranging from 25 to 35, which are of a comparable size with the dispatching territories used in the test cases presented in Section **??**, the expected numbers of variables and constraints are above 300,000 for a given planning horizon of 8-12 hours.

On average, experimental SDTP-scenarios were solved by Cplex in $\sim$90 minutes with a 10% mipgap. Since holistically stitching the entire railway network with multiple dispatching territories requires solving each single dispatching territory multiple times, it would be practically impossible to utilize a MIP model. Hence, we chose to use an Oracle which utilizes a heuristic and yields results as good as the mipgap 2-5% solutions under 10 seconds.