AN EVENT BASED OPTIMISATION MODEL FOR TRAIN SCHEDULING

Sam Pyke and Erhan Kozan

School of Mathematical Sciences, Queensland University of Technology,

GPO Box 2434 Brisbane Qld 4001 Australia

sj.pyke@student.qut.edu.au, e.kozan@qut.edu.au

ABSTRACT

In this paper, a new model is developed for scheduling railway operations. Due to the combinatorial nature of the problem, which results in the difficulty increasing exponentially with the size of the problem, previous efforts at solving this problem have tended to focus on simple cases - mostly sections of track with no branch-lines - although they may consist of a mixture of single- and double-line track. Previous models have also tended to ignore the need for acceleration and deceleration, in order to express the model as a relatively simple (but still difficult to solve) mixed-integer linear program. A more general model of rail operations, including branch-lines, acceleration and deceleration, was constructed. The model can be solved exactly by branch-and-bound, but for large problems this is too slow to be suitable for realtime applications. Various heuristic techniques are to be examined to find a method for rapidly generating good (if not perfect) solutions.

Key Words: Railway Engineering and Technologies, Scheduling

1. INTRODUCTION

This paper describes a new model for train scheduling, including the effects of acceleration, deceleration and train length, as well as allowing branch lines and taking estimated risk of delay into consideration. Acceleration and deceleration introduce nonlinearities, but allowing speed changes to be instantaneous as previous models do makes the model linear. Train length and branch lines are handled without complicating the model, by considering each train movement as a sequence of events. The risk of delay used in the model is only an estimate, since exact calculations would be extremely time-consuming, and as a result the model would be effectively unsolvable.

Section 2 gives a brief summary of past work on modelling train scheduling and estimating schedule reliability. Sections 3 and 4 describe the model. Section 5 contains some further remarks on the model.

2. PAST WORK

2.1. Train Scheduling Models

Much past work on train scheduling models (see for example Cai *et al* (1998), Cheng (1998), Higgins (1996), Higgins *et al* (1996), Şahin (1999) and Zhou and Zhong (2004)) has focussed on single corridors. There have also been models that do allow branch lines, with various levels of detail, such as the following. Adenso-Díaz *et al* (1999) considered the meet-pass plans as given, and considered how to recover from delays by cancelling services or reassigning trains. This method of resolving timetables is not universally applicable, since cancelling services may conflict with a particular railway's policy, and reassigning trains is only possible if there are spare train units available in the right place. Kraay and Harker (1995) did not consider the exact meet-pass plan, assuming that the amount of delay a train will suffer when passing or overtaking another between two given points, which may be stations or junctions, can be determined without considering where in that area the trains meet. Carey (1994a, 1994b) produced a more detailed model, but it could be considered too detailed in that it determines the exact path taken by a train, right down to which side of a duplicated track it uses—and the solution technique cannot avoid exploring all combinations of such options. Chiang *et al* (1998) developed an expert-system based scheduler for single corridors, and described how to extend it to general train networks. Walker *et al* (2004) handles meet-pass planning in a way that, while assuming trains' routes are fixed, makes no assumptions about the topology of the rail network, except that only single-line tracks are considered; multi-line tracks may occur, but they are implicitly assumed never to be the subject of a conflict. Stations and crossing loops are also implicitly assumed to have infinite capacity.

The detailed models above (ie the above models except Adenso-Díaz *et al* (1999) and Kraay and Harker (1995)) have several weaknesses in common:

- Except for Zhou and Zhong (2004), and possibly Chiang *et al* (1998), they ignore acceleration and deceleration, which can result in inaccuracies of several minutes for some trains, such as coal trains. These inaccuracies have the potential to accumulate, ruining the entire schedule.
- They treat trains as points without length, assuming that any pair of trains can cross at any crossing loop. In fact, some trains are longer than some crossing loops, and a pair of trains can only cross at a crossing loop longer than at least one of them.
- Of all the above models (including the two less detailed ones) only Higgins (1996) even attempts to consider the reliability of the schedule.

Other train scheduling models have been developed which are sufficiently similar to some of the above that they need not be treated separately. For example, Ghoseiri *et al* (2004) uses a model with similar constraints to Carey (1994a, 1994b) crossed with Higgins et al (1996), but uses a multi-objective formulation that includes fuel consumption. It shares the three weaknesses listed above.

2.2. Schedule Reliability Estimation

Higgins (1996) broke exogenous delays into three types, and derived estimates of the knock-on delays resulting from each type of exogenous delay. However, the resulting delay estimate is complicated and not well-suited to inclusion in a rapid optimisation model. Carey (1999) examined reliability estimates using and not using probabilities. The estimates using probabilities must be calculated recursively, with an integration at each step, and are therefore not suitable for inclusion in an optimisation model. The estimates not using probabilities are various measures of headways that are difficult to apply to trains travelling in a mixture of directions, and a heuristic based on the number of conflicts, which is not applicable to an optimisation model (since it is not significantly affected by the optimisation variables). Chen and Harker (1990) put forward a reliability model that assumes that delays are normally distributed—which can only be anywhere near true if on-time running is rare. Various older models assume that trains are distributed randomly—which can only be true if there is no schedule, or the schedule has completely disintegrated.

3. NOTATION

3.1. Parameters

J = the set of train movements from one scheduled stop to the next
E_j = the set of events on movement j
$\underline{e}_j$ = the initial event of movement j
$\bar{e}_j$ = the final event of movement j
$e+1$ = the event immediately following event e on the same movement
$e-1$ = the event immediately preceding event e on the same movement
P_{je} = the set of events in E_j preceding event e
D_j = the scheduled departure time for movement j
A_j = the scheduled arrival time for movement j
f_j = the penalty function for the tardiness of movement j (to be used as a component of the overall objective function)
w_e^k = the weighting function for knock-on delay resulting from an exogenous delay at event e
C_j = the set of pairs of connections (j',t), where t is the amount of time that must elapse after the completion of movement j' before movement j begins.
U_j = the initial velocity of movement j
B = the set of track sections, from one signal (or equivalent) to the next (equivalent to a safeworking block)
Φ_b^s = the number of vacant tracks in section b at the start
Φ_b^e = the number of vacant tracks in section b at the end
I_b = the set of events whereby a train enters section b
O_b = the set of events whereby a train leaves section b
δ_{be} = the minimum time that can elapse between an event on section b and event e if the event on section b is to affect event e
$\overline{V}_{ee'}(t,v)$, $\underline{V}_{ee'}(t,v)$ = upper and lower bounds on $V_{e'}$ where $t=T_e$ and $v=V_e$
$\overline{T}_{e+1}(t,v,v')$, $\underline{T}_{e+1}(t,v,v')$ = upper and lower bounds on T_{e+1} where $t=T_e$, $v=V_e$ and $v'=V_{e+1}$
M = a really big number

Each train service consists of a series of one or more movements, ideally from one station to the next, although movements as considered in the model may begin or end at the border of the region considered by the scheduler. The events of a movement include entries to and exits from sections, as well as the beginning and end of a movement. A train, since it has nonzero length, must, whether or not the safeworking blocks actually overlap, enter one section before leaving the previous one. A train that is longer than a particular section that it must cross (eg a long train at a short crossing loop) will temporarily occupy three sections simultaneously (which, in the crossing loop example, limits its ability to use the loop).

In many cases, $f_j(x)=w_j x$ where w_j is a weighting factor expressing how important it is that movement j is completed on time. Even if it isn't, it should be differentiable, and close to linear on scales similar to the size of likely delays. w_e^k will typically be the pdf of exogenous delay to event e (or some approximation thereto), possibly multiplied by some constant greater than 1 to indicate that the scheduling should be risk-averse.

C_j connects movements that must be performed consecutively by a single train, as well as movements that must wait for passengers or cargo to be transferred. U_j will usually be zero, but may be non-zero for movements coming from outside the area considered. δ_{be} depends on the speed of the signalling system. $\overline{V}_{ee'}(t,v)$, $\underline{V}_{ee'}(t,v)$, $\overline{T}_{e+1}(t,v,v')$ and $\underline{T}_{e+1}(t,v,v')$ depend on the locations of the events, the train type, the curvature and slope of the track and possibly other operational restrictions (such as safety regulations and noise restrictions).

3.2. Variables

T_e = the time at which event e occurs.

L_j = the tardiness of journey j

V_e = the speed at which event e occurs, ie the speed at which the train carrying out event e is moving at T_e

V'_e = an estimate of the highest speed that event e can occur at (in order to catch up after a delay). It is exact if the acceleration and deceleration limits do not depend on time and the delay to the train occurred no later than the last time it was stationary before event e. Even when it is not exact, it is always at least V_e.

$k_{ee'}(\tau)$ = an estimate of the delay to event e' resulting from an exogenous delay of τ time-units to event e

$$N_{ee'} = \begin{cases} 1 & \text{if event } e' \text{ (an entry to some section) is the next event after} \\ & \text{event } e \text{ (an exit from the same section) to use the same track} \\ 0 & \text{otherwise} \end{cases}$$

$$\phi_e = \begin{cases} 1 & \text{if event } e \text{ begins the first occupation of a particular track} \\ 0 & \text{otherwise} \end{cases}$$

$$\lambda_e = \begin{cases} 1 & \text{if event } e \text{ ends the last occupation of a particular track} \\ 0 & \text{otherwise} \end{cases}$$

$$\beta_{ee'} = \begin{cases} 1 & \text{if event } e \text{ occurs at least } \delta_{be'} \text{ before event } e' \text{, where } e \in O_b \\ 0 & \text{otherwise} \end{cases}$$

The descriptions of $N_{ee'}$, ϕ_e and λ_e correspond to one way of assigning tracks. Other track assignments may be possible for the same schedule.

The integer part of the model is not as huge as it appears, since:

- $N_{ee'}$ only applies where $e \in O_b$ and $e' \in I_b$ for some b,
- ϕ_e only applies where $e \in I_b$ for some b,
- λ_e only applies where $e \in O_b$ for some b,
- many of the $\beta_{ee'}$ s can be fixed and
- the $\beta_{ee'}$ s not appearing in Equations 10 and 11 can be ignored.

4. THE MODEL

The only non-linearities are in Equations 4–9, 11 and 24. Allowing speed changes to be instantaneous (as most previous models assume) allows equation 11 to be eliminated and makes the others linear.

4.1. Objective Function

$$\sum_{j\in J} f_j(L_j)+\sum_{j\in J}\sum_{e\in E_j}\int_0^{\infty} w_e^k(\tau)\sum_{j'\in J} f_{j'}(L_{j'})k_{e\bar{e}_{j'}}(\tau)\,d\tau \tag{1}$$

The first term is the total weighted lateness if everything runs according to the schedule determined by solving the model. The second term is the estimated risk of delay, which is the sum over each e and τ of the delay resulting from an exogenous delay of τ time-units to event e, weighted by $w_e^k(\tau)$. For simplicity the model considers only one exogenous delay at a time. The integration can be replaced by a summation by taking a suitably chosen sample of values of τ.

4.2. Physical Constraints

Each train's initial velocity is given:

$$V_{\underline{e}_j} = V'_{e\underline{e}_j} = U_j \quad \forall j, j' \in J, e \in E_{j'} \tag{2}$$

If a train is already moving, it is moving now:

$$T_{\underline{e}_j} = D_j \quad \forall j \in J, U_j \neq 0 \tag{3}$$

A train's acceleration, deceleration and top speed are limited:

$$V_{e+1} \le \overline{V}_{e(e+1)}(T_e, V_e) \forall j \in J, e \in E_j, e \neq \bar{e}_j \tag{4}$$

$$V'_{e+1} \le \overline{V}_{e(e+1)}(T_e, V'_e) \forall j \in J, e \in E_j, e \neq \bar{e}_j \tag{5}$$

$$V_{e+1} \ge \underline{V}_{e(e+1)}(T_e, V_e) \forall j \in J, e \in E_j, e \neq \bar{e}_j \tag{6}$$

$$V'_{e+1} \ge \underline{V}_{e(e+1)}(T_e, V'_e) \forall j \in J, e \in E_j, e \neq \bar{e}_j \tag{7}$$

$$T_{e+1} \le \overline{T}_{e+1}(T_e, V_e, V_{e+1}) \forall j \in J, e \in E_j, e \neq \bar{e}_j \tag{8}$$

$$T_{e+1} \ge \underline{T}_{e+1}(T_e, V_e, V_{e+1}) \forall j \in J, e \in E_j, e \neq \bar{e}_j \tag{9}$$

4.3. Safeworking Constraints

A train must not enter a section of track until the previous one is out of the way:

$$\beta_{ee'} \ge N_{ee'} \quad \forall b \in B, e \in O_b, e' \in I_b \tag{10}$$

A train must be ready to stop before any section of track that isn't clear yet:

$$\underline{V}_{e+1,e'}(T_e, V'_e) \le M(1 - N_{e''e'}) + M\beta_{e''e} \quad \forall j \in J, b \in B, e' \in E_j \cap I_b, e \in P_{je'}, e'' \in O_b \tag{11}$$

In order to enforce the above constraint on V, we need only enforce the meaning of V':

$$V_e \le V'_e \quad \forall j \in J, e \in E_j \tag{12}$$

In any section, there must be at most one train per track:

$$\sum_{e\in O_b} N_{ee'} + \phi_{e'} = 1 \forall b \in B, e' \in I_b \tag{13}$$

$$\sum_{e'\in I_b} N_{ee'} + \lambda_e = 1 \forall b \in B, e \in O_b \tag{14}$$

$$\sum_{e\in I_b} \phi_e \le \Phi_b^s \forall b \in B \tag{15}$$

$$\sum_{e\in O_b} \lambda_e \le \Phi_b^e \forall b \in B \tag{16}$$

Calculate β:

$$T_e + \delta_{be'} \le T_{e'} + M(1 - \beta_{ee'}) \quad \forall b \in B, e \in O_b, j \in J, e' \in E_j \tag{17}$$

4.4. Timetable Constraints

Trains must not depart early, since we don't want to leave passengers or cargo behind:

$$T_{\underline{e}_j} \geq D_j \quad \forall j \in J, U_j = 0 \tag{18}$$

A train must remain at each station for the required dwell time, and must wait for connecting trains:

$$T_{\underline{e}_j} \geq T_{\bar{e}_{j'}} + t \quad \forall j \in J, (j', t) \in C_j \tag{19}$$

Each train must stop at its destination:

$$V_{\bar{e}_j} = V'_{\bar{e}_j} = 0 \quad \forall j \in J \tag{20}$$

Calculate lateness:

$$L_j \geq T_{\bar{e}_j} - A_j \quad \forall j \in J \tag{21}$$

4.5. Delay Propagation

If an event is delayed, it is delayed:

$$k_{ee}(\tau) = \tau \quad \forall j \in J, e \in E_j, \tau \geq 0 \tag{22}$$

If a train is late arriving at a station, it, and any connecting trains, may be late leaving the station:

$$k_{e\underline{e}_j}(\tau) \geq k_{e\bar{e}_{j'}}(\tau) - T_{\underline{e}_j} + T_{\bar{e}_{j'}} + t \quad \forall j, j'' \in J, e \in E_{j''}, (j', t) \in C_j \tag{23}$$

Once a train is delayed, it stays delayed until it catches up:

$$k_{e(e'+1)}(\tau) \geq k_{ee'}(\tau) - T_{e'+1} + \underline{T}_{e'+1}(T_{e'}, V'_{e'}, V'_{e'+1}) \quad \forall j, j' \in J, e \in E_j, e' \in E_{j'}, e' \neq \bar{e}_{j'} \tag{24}$$

If a train is late leaving a section of track, another may be delayed waiting for the track to be free:

$$k_{ee'}(\tau) \geq k_{ee''}(\tau) - T_{e'} + T_{e''} + \delta_{be'} - M(1 - N_{e''e'}) \quad \forall j \in J, e \in E_j, b \in B, e' \in I_b, e'' \in O_b \tag{25}$$

4.6. Alternative Representation of Safeworking Constraints

The safeworking constraints can be written in an alternative form that produces the same results. For example, by matching each section entrance and exit that begin and end the section occupation, an extension of the system in Higgins *et al* (1996) can be used, whereby Equations 10 and 13–16 are replaced by Equations 26 and 27, and $N_{e''e'}$ is replaced with $\beta_{e''e'}$ in Equations 11 and 25.

$$\beta_{o_c i_{c'}} + \beta_{o_{c'} i_c} + s_{cc'} = 1 \forall b \in B; c, c' \in K_b \tag{26}$$

$$\sum_{c' \in C_b} s_{cc'} \leq \kappa_b \forall b \in B; c \in K_b \tag{27}$$

K_b is the set of occupations of section b, i_c and o_c are the events beginning and ending the occupation c, κ_b is the number of lanes in section b, and $s_{cc'}$ is a 0-1 variable which is 1 if occupations c and c' occur simultaneously. Note that $s_{cc} = 1$ for all c, and that $s_{cc'} = 0 \forall b \in B; \kappa_b = 1; c, c' \in K_b; c \neq c'$, so the simultaneity variables are, as one might expect, only relevant on multi-lane blocks. If c is at the beginning or end of the period considered, then i_c or o_c, respectively, will be non-existent and the corresponding term should be absent from the instances of Equation 26 that c occurs in.

5. EXPERIMENTAL RESULTS

Both representations of the model have been tested on an example problem, with eighteen train services, carried out by five train units. The services consist of a total of twenty-eight movements, on a rail network with eleven sections, three of which are on a branch line. There is a mixture of single- and double-track sections, four stations, of which three have duplicated track, and a crossing loop which is shorter than three of the train units. In Figure 1, the block boundaries are marked with vertical lines, and the stations with rectangles. The boundary at the right marks the edge of the scheduler's zone of control; four of the trains come from and return to an independently operated train network to the east.

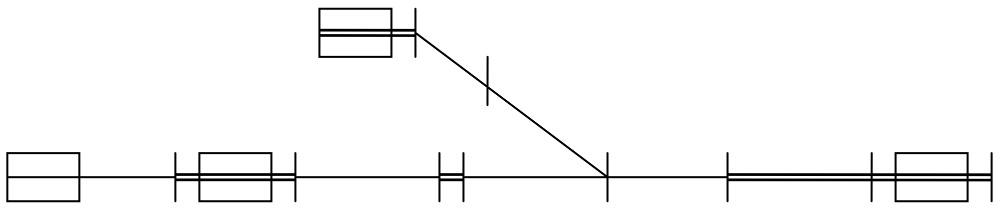

Figure 1. Layout of the Train Network

Solution of both representations of the model was attempted using a general-purpose MILP solver (ILOG CPLEX 8.0.0) on a 2.4GHz Intel Xeon processor, with 512MB of RAM and just over a gigabyte of swap space. Other programs were in memory at the time, but more than half of the memory was available, even without swapping to disk.

The first representation caused CPLEX to run out of memory, but the alternative representation in Section 4.6 was solved in approximately a minute (it was run several times, and actual running times varied by tens of seconds). The particular example chosen turned out not to require any trains to be late.

6. CONCLUSIONS AND FUTURE STUDY

Since the only indication that trains in this model are travelling on the same track is that they enter and leave the same sections, the event sequences can easily reflect any track topology, so long as the trains' routes are fixed. Choice of tracks of a multi-track section is handled by assuming that it is irrelevant, which is true for some train networks, and near enough for others.

The difference in solvability between the two representations is a striking example of the benefits that can be gained by reformulating a model. Presumably the difference results from the different ways that fixing the integer variables divides the solution space; there is reason to suppose that the second representation results in more even divisions, and that the first excessively distinguishes similar schedules.

It seems likely that a special-purpose branch-and-bound algorithm would solve the problem faster than a general-purpose MILP branch-and-bound solver such as the one used. However, the problem is NP-complete, and large instances cannot be solved exactly in a reasonable amount of time. This is especially a problem for dynamic train scheduling (ie rescheduling to minimise the adverse effect of an unforseen event), in which a good plan now is better than a perfect plan tomorrow. Heuristic techniques for rapidly finding good solutions are being investigated as part of the research program this paper arises out of.

REFERENCES

Adenso-Díaz, B., Oliva González, M. and González-Torre, P. (1999), On-Line Timetable Re-Scheduling in Regional Train Services, *Transportation Research*, **33B**, 387–398.

Cai, X., Goh, C. J. and Mees, A. I. (1998), Greedy Heuristics for Rapid Scheduling of Trains on a Single Track, *IIE Transactions*, **30**, 481–493.

Carey, M. (1994a), A Model and Strategy for Train Pathing with Choice of Lines, Platforms and Routes, *Transportation Research*, **28B**, 333–353.

Carey, M. (1994b), Extending a Train Pathing Model from One-Way to Two-Way Track, *Transportation Research*, **28B**, 395–400.

Carey, M. (1999), Ex Ante Heuristic Measures of Schedule Reliability, *Transportation Research*, **33B**, 473–494.

Chen, B. and Harker, P. T. (1990), Two Moments Estimation of the Delay on Single-Track Rail Lines with Scheduled Traffic, *Transportation Science*, **24**, 261–275.

Cheng, Y. (1998), Hybrid Simulation for Resolving Resource Conflicts in Train Traffic Rescheduling, *Computers in Industry*, **35**, 233–246.

Chiang, T.-W., Hau, H.-Y., Chiang, H.-M., Ko, S.-Y. and Hsieh, C.-H. (1998), Knowledge-Based System for Railway Scheduling, *Data & Knowledge Engineering*, **27**, 289–312.

Ghoseiri, K., Szidarovsky, F. and Asgharpour, M. J. (2004), A Multi-Objective Train Scheduling Model and Solution, *Transportation Research*, **38B**, 927–952.

Higgins, A. (1996), *Optimisation of Train Schedules to Minimise Transit Time and Maximise Reliability*, PhD thesis, School of Mathematics/School of Civil Engineering, Queensland University of Technology.

Higgins, A., Kozan, E. and Ferreira, L. (1996), Optimal Scheduling of Trains on a Single Line Track, *Transportation Research*, **30B**, 147–161.

Kraay, D. R. and Harker, P. T. (1995), Real-Time Scheduling of Freight Railroads, *Transportation Research*, **29B**, 213–229.

Şahin, İ. (1999), Railway Traffic Control and Train Scheduling based on Inter-Train Conflict Management, *Transportation Research*, **33B**, 511–534.

Walker, C. G. C. G., Snowdon, J. N. and Ryan, D. M. (2004), Simultaneous Disruption Recovery of a Train Timetable and Crew Roster in Real Time, *Computers & Operations Research*, in press.

Zhou, X. and Zhong, M. (2004), Bicriteria Train Scheduling for High-Speed Passenger Railroad Planning Applications, *European Journal of Operational Research*, in press.