TIME EFFICIENT GREEDY STRATEGY FOR SCHEDULING TRAINS ON A LINE

J. Medanic and M.J. Dorfman

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Abstract: Scheduling trains on a single railway line is an issue in its own right, and is a building block for scheduling trains in railway networks. A local, state-dependent, travel advance strategy combined with a discrete event model of a railway line represent a more efficient way of approaching the scheduling problem than nonlinear programming approaches used in the past. The approach also eliminates a deficiency of nonlinear programming formulations, which produce a programmed schedule that cannot be applied if any perturbation in train operations occurs.

Keywords: Discrete event dynamic systems, Scheduling, Algorithms, Train control

1. INTRODUCTION

Railroad transportation in the US is a major factor in freight transportation. While Japan and Western Europe have a dense and well-developed passenger system the US boasts the most developed freight transportation system in the world. The freight railways are continuously increasing their reliance on modern technology (e.g. introduction of Electronically Controlled Pneumatic brakes and Positive Train Control) to enable more efficient train operation. At the same time, despite the fact that railways carry about 40% of all freight in the US, track miles in use have steadily declined, many of the existing railway corridors have single-track lines, and there are no plans for adding a second track to existing corridors. In addition, rail freight transportation in the US is based on diesel traction, and the cost effectiveness of transportation of freight by rail is becoming more sensitive to the increasing costs of fossil fuel, a cost that will only increase in the near future. More efficient train operation, and specifically energy- and time-efficient scheduling, as well as fast rescheduling of trains, can significantly contribute to the cost effective operation of freight traffic.

2. SCHEDULING TRAINS ON A LINE

Three main approaches have been pursued in solving scheduling problems, the choice depending on the characteristic features of the scheduling problem: (i) linear or nonlinear programming formulations, (ii) network flow formulations, and (iii) dynamic system formulations. The main feature of the problem is that trains traveling in opposite directions can meet and pass (M&P) each other, and trains traveling in the same direction can meet and overtake (M&O) each other only at sidings and stations (referred to here as M&P points). A line may contain some double track sections (or long sidings) which allow additional M&P and M&O opportunities, but otherwise the characteristics of the problem remain the same. A railway network scheduling problem essentially retains these characteristic features with cross-over and merge points.

The scheduling problem with obstruction constraints is not the most appealing type of problem for a programming approach because it involves a huge number of precedence conditions (times of arrival and departures of trains at M&P points) and logical variables to obtain a valid formulation (e.g. Higgins et al., 1995). Branch and bound algorithms for integer programming problems have been used to solve the problem (Kraft, 1987). The main drawback, in addition to the significant computational effort (which often requires that part of the scheduling problem be simplified, such as estimating future delays, as opposed to computing the actual delays), is that the obtained schedule is valid only if no perturbation occurs in its realization.

The problem does not fit well into network flow formulations widely employed in communication and computer networks, and also in some job scheduling problems in manufacturing. Such problems are formulated in terms of five basic characteristics at each node of a network: the arrival pattern, the service pattern, the number of parallel servers, the...
The problem formulation used here differs from those used in the programming approaches in that the departure times and train velocities in sections are assumed fixed (as opposed to belonging to pre-defined admissible ranges), and the stop times at M&P points, arrival times at destinations, and the complete schedule are obtained by applying the TAS and solving the discrete event dynamics.

The approach, applied here to develop a greedy TAS to be described, provides more information and the solution has features different than solutions obtained using nonlinear programming approaches. First, for given departure times and velocities, it determines a complete schedule. Second, the greedy TAS can be used to develop schedules for perturbed cases such as when a particular train is off its schedule, and the scheduling of all trains must be modified, or when a temporary speed restriction requires a change in schedule, when an additional train must be introduced into a given schedule, etc. Third, the computational effort is extremely moderate, comparable to solving for the time trajectories of a dynamic systems of order \( N=N_1+N_2 \), where \( N_1 \) is the number of trains traveling in one direction and \( N_2 \) the number of trains traveling in the opposite direction. (Although the algorithm, and the associated software, contains many logical “if” types of pieces of code, beyond the simple calculation of train advances in each discrete event step, unique paths through such logical statements do not increase the computation time as is the case in the programming formulations when all constrains must be checked for feasibility.) Fourth, the greedy TAS has been shown to have highly desirable characteristics in relation to scheduling trains on a single line. If nominal train velocities by sections are the maximum allowed velocities (due to track, infrastructure, or train restrictions) the greedy TAS provides time efficient advance of all trains; using the ratio of unobstructed times to unobstructed times required for all trains to clear the line as a performance criterion, denoted by \( \eta \), the schedule developed by the TAS produces schedules with \( \eta \) in the range \([0.95 – 0.99]\) for a variety of scheduling problems. Moreover, the greedy TAS easily modifies into a strategy with optimal pacing velocities while maintaining the above time efficiency ratio. Thus, it can provide an energy efficient solution with optimal pacing velocities in the spirit of the prevailing philosophy in train scheduling.
scheduling. Finally, the greedy TAS exposes the local nature of decision-making that suffices in most train encounters, and so offers realistic extension of the approach to scheduling trains in a railway network.

The train model used in scheduling studies assumes that the velocity is constant on sections of the line, and we accept this assumption, although the variable train velocity case is discussed as well. In addition the following assumptions are made: (i) The route is fixed and defined by the vector $x_d$, (ii) Velocities of all trains in all sections of the route are fixed, and given by the matrices $V_L$ and $V_R$, respectively; (i.e. the element $V_L(i,m)$ is the velocity of train $i$ traveling from O to D in section $m$, and the element $V_R(j,n)$ is the velocity of train $j$ traveling from D to O in section $n$); (iii) The times of origin of the trains are given by the vectors $T_{OL}$ and $T_{OR}$, and the arrival times are free, and depend on the train advance strategy; (iv) The minimal headways of trains are defined by the vectors $d_i$ and $d_j$ (i.e. $d_i(i)$ defines the minimum distance between train $i$ and any train ahead of it).

Assuming constant velocities, the model of system dynamics may be written in the form

$$
x(k+1) = x(k) + \Delta t \sum_{j=1}^{N} v_L(x(k), y(k)), \quad y(0) = x_0
$$

$$
y(k+1) = y(k) + \Delta t \sum_{j=1}^{N} v_R(x(k), y(k)), \quad y(0) = y_0
$$

(1)

where the train positions form the state, and the time periods $\Delta t$ in the discretization vary and depend on the state and train velocities, with the constant velocity of a train depending on the section of the line the train is currently traversing. The time intervals are triggered by the arrival of trains at stations or sidings. The trains reach M&P points asynchronously, and this leads to a discrete event system (DES).

The train advance strategy proposed here will be referred to as a greedy TAS because it is locally optimal and depends on local information. The advance of train $i$, moving in the O to D direction, and train $j$, moving in the same direction and any train $j$ moving in the opposite direction and immediately ahead of train $i$, as will be defined (and vice versa for a train $j$ moving in the D to O direction). The main components of the greedy TAS are:

(a) Determination of the next discrete event (the train which will first reach an M&P point, and the required time interval $t_{next}$).

(b) Resolution of the M&P and/or M&O events at this M&P point, and possibly at other M&P points where a train is stationed-at the current discrete event, and

(c) Development of the rules for a simple M&P, simple M&O, or a combined M&P with M&O event.

Let the vector $x_d$ with $k+1$ component delineate lengths of sections of the line, with $x_d(1) = 0$ and $x_d(k+1)-x_d(k)$ the length of section $k$. Indexing by $L$ trains traveling from O to D, and by $R$ trains traveling from D to O, two vector variables, $S_L$ and $P_L$, will be associated with trains moving from O to D, and two vector variables, $S_R$ and $P_R$, will be associated with trains moving from D to O. Variables $S_L$, $S_R$ are used to characterize the velocities of trains while in a section, variable $P_L$, $P_R$ are used for a train at an M&P point. Thus, $P_L(i) = n$ implies $x_d(P_L(i)) = x_d(n)$ and identifies the train $i$ as being at M&P point $n$, etc.

Given a state $(x(k), y(k))$ at some discrete event (DE) $k$, the time to next M&P for each train is computed, from

$$z(i) = \frac{|x(i,k) - x_d(P_L(i,k))|}{V_L(i,S_L(i,k))}, \quad i = 1, 2, ..., N_L$$

$$w(i) = \frac{|y(i,k) - x_d(P_R(i,k))|}{V_R(i,S_R(i,k))}, \quad j = 1, 2, ..., N_R$$

and adjusted by eliminating the associated components of $z(i)$ or $w(i)$ if a slower train is obstructing a faster train in reaching first an M&P point. Here

$$S_L(i) = m \text{ if } x(L,m) \in [x_d(m),x_d(m+1))$$

$$P_L(i) = m \text{ if } x(L,m) \in [x_d(m),x_d(m+1))$$

$$S_R(j) = n \text{ if } y(R,n) \in [x_d(n),x_d(n+1))$$

$$P_R(j) = n \text{ if } y(R,n) \in [x_d(n),x_d(n+1))$$

$$V_L(i,k) = \text{ velocity of train } i \text{ in section } k$$

$$V_R(j,k) = \text{ velocity of train } j \text{ in section } k$$

Given an arbitrary vector $\Gamma$ let the two arguments $\alpha, \beta$ in the operation $[\alpha, \beta] = \min(\Gamma)$ denote the minimal component and the lexicographical order of that component in $G$. Then, given the vectors $z$ and $w$, let

$$z_{min} \leq z_{min} \leq w_{min}$$

characterize the train $(i_{min}, j_{min})$ to first reach the next M&P point and the minimum time required $z_{min} \leq w_{min}$, at the current discrete event. When trains are not in the vicinity of each other all trains will advance along the line for the duration of the time interval

$$t_{next} = \min(z_{min}, w_{min})$$

(4)

at which time the next DE occurs (because train $i_{min}$ or $j_{min}$, as the case may be, reaches an M&P point, referred to as the focal M&P for that DE).

Concerning the service rules, referred to here as rules of engagement, a strategy must be defined for an M&P event, an M&O event and a combined M&P/M&O event involving three trains. four trains cannot pass each other if they are in vicinity of each other at any one DE, and this option is excluded by the defined rules of engagement at prior discrete events.
If two trains traveling in the same direction are in the vicinity of each other, there exists a case when the dtnext must be computed differently because of a slower train obstructing a faster train. There are two subcases to consider: (a) train i-I is the first to reach an M&P point, in which case \( dtnext = z_{(i-1)} \), or (b) there is some other train that reaches some other M&P point first, and \( dtnext \neq z_{(i-1)} \). In subcase (a) trains i-I and I are both advanced to \( P_I(i-I) = m+1 \). At this moment, as far as the train advance strategy is concerned, an overtake has occurred, and the order of trains moving in the O to D direction is changed, train i becoming i-1, and train i-I becoming train I. In subcase (b) some other train, which first reaches an M&P point, defines dtnext and the next implementable DE; train i-I is advanced towards the M&P point, and train i is advanced as close to it as its minimum headway allows.

The greedy TAS decomposes all M&P events into eleven distinct M&P, M&O and combined M&P/M&O events and defines for each train in vicinity of the focal M&P event which train is stopped and which train advances in the current DE.

Concerning a M&P event, two trains, say i and j, in the vicinity of each other can be in any one of the following situations:

(i) i is at M&P point m (with PL(i) = m), j is in section m (with SR(j) = m)
(ii) i is in a section m (SL(i) = m), j is at M&P point \( m+1 \) \( (PR(j) = m+1) \)
(iii) i is at M&P point m (PL(i) = m), j is at M&P point \( m+1 \) \( (PR(j) = m+1) \)
(iv) i is in a section m (SL(i) = m), j is in section \( m+1 \), \( (SR(j) = m+1) \).

Concerning M&O event between two trains i, i-I moving from O to D, and in the vicinity of each other, the following cases are to be distinguished:

(i) train i-I is faster, and is either traveling in Section \( S_i(i-I) = m \) or is at the M&P point \( P_i(i-I) = m \) while train i is in section \( S_i(i) = m \);
(ii) train i is faster and is traveling on section \( S_i(i) = m \), while train i-I is at the M&P point \( P_i(i-I) = m+1 \);
(iii) train i is faster and is traveling in section \( S_i(i) = m \), and train i-I is also traveling ahead of it in section \( S_i(i-I) = m \).

Case (ii) is the only case where an overtake can occur at the considered DE. Whether an overtake will take place or not is determined by the speed at which the two trains travel. Since train i-I is at an M&P point, and is slower, the time it needs to reach the next M&P point is longer than for train i. In the greedy strategy an overtake will occur if

\[ V_i(i,S_i(i)) > V_i(i-I,S_i(i-I)). \] (5)

If (5) holds train i-I will be held at PL(i-I) = m, and train i will be advanced to PL(i) = m. In the opposite case, train i-I will be advanced, except if there is also a train j traveling in the opposite direction in the vicinity giving rise to a simultaneous M&P event, to be discussed below.

The different situations involving three trains that may be encountered in case of a combined M&P/M&O event, considered from the point of view of a train i traveling from O to D, and meeting two train traveling from D to O, are:

(i) train i is in section SL(i) = n, and the trains j-1, j are in section \( PR(j) = n+1 \);
(ii) train i is at the M&P point \( P_i(i) = n+1 \), and the trains j-1, j are in section \( PR(j-1) = P_j(j) = n+1 \);
(iii) train i and train j-1 are at the M&P point n, train j is traveling in section \( n+1 \);
(iv) train i is in section SL(i) = n, train j-I is at the M&P point n, and train j is in section \( n+1 \).

A rule of engagement is defined for each case (omitted here due to space limitations) and the totality of these rules form the greedy TAS. This is a local strategy because only trains in the vicinity of each other enter into the decision affecting which train will advance, and which will be stopped at an M&P point. Its general form is

\[ u_i(k) = u_i[x(i,k),x(i-1,k),y(jv, k),y(jv, v)] \]
\[ v_j(k) = v_j[x(i, v, k),y(jv, k),y(j, k, v),y(jv, v)] \] (6)

where \( j_v, j-1 \) and \( k_v, k-1 \) denote trains (if any) in the vicinity of train i, or train j, respectively. It application requires that each train obtain information on the front in front of it traveling in the same direction, as well as closest train(s) approaching it from the opposite direction. If all train operators adhere to the strategy and there is a perturbation in the schedule of any particular train, one can apply the strategy to efficiently re-compute the remaining schedule from the new state, and apply it.

4. TIME-EFFICIENT PERFORMANCE OF THE GREEDY STRATEGY

A number of performance measures were used to assess the time-keeping features of the proposed strategy: (i) the time to clear the line, (ii) the total delay of all trains, and (iii) the maximum delay. The time to clear the line of all trains, is defined as

\[ J_t = t_{N_a} - t_{t_d} \] (7)

where \( t_{t_d} \) is the time of departure of the earliest train on the schedule, and \( t_{N_a} \) is the time of arrival of the latest train on the schedule. A characteristic of this criterion is that the total time to clear the line if all trains travel unobstructed is the minimum possible value of \( J_t \), denoted by \( J^*_t \). (This correspond to the
Given departure times and velocities one can compute \( t_{N_a}^f \), the time of arrival of the latest train and \( J_1^f = t_{N_a}^f - t_{1d}^f \). The **efficiency ratio**

\[
\eta = \frac{t_{N_a}^{ob} - t_{1d}^{ob}}{t_{N_a}^{f} - t_{1d}^{f}}
\]  

(8)
is taken as the measure of the time-efficiency of the schedule obtained by a TAS. The superscript “ob” stands for obstructed time, the time of arrival of the last train as computed from the greedy schedule. The **total delay** criterion is defined by

\[
J_2 = \sum_{i=1}^{N} (T_{i_a}^{ob} - T_{i_a}^{f})
\]

(9)

where \( T_{i_a}^{ob} \) is the time of arrival of train \( i \) as obtained by a greedy TAS while \( T_{i_a}^{f} \) is the time of arrival of train \( i \) with unobstructed travel. The maximum delay criterion is defined by

\[
J_3 = \max_{i} (T_{i_a}^{ob} - T_{i_a}^{f})
\]

(10)

Performance of the greedy TAS was analyzed on numerous examples of scheduling trains within a 24 h period. The main conclusion is that with number of trains traversing a single line in a day, from each direction, of the order of 2 per hour the greedy strategy easily determines a schedule without encountering a deadlock. Moreover, as the number of trains is increased the efficiency ratio remains remarkably constant. The greedy TAS consistently produces schedules with \( \eta \) in the range 0.95-0.99, with \( \eta = 1 \) the minimal possible value.

**Example 1.** We will illustrate the quality of the results that are obtained on a hypothetical study of the capacity of a line to a specific composition of trains traversing it. In brief, a line with 11 single track sections was defined with total length of 210 [mi] and with different maximum velocities, varying between 50-90 [mi/hr] in each section, but the same for all trains. The headways were arbitrarily set at 0.5 [mi] for all trains. The same number of trains was assumed to depart from each end of the line, with departure times of trains approximately uniformly distributed over a 24 h period. The number of trains was then increased from \( N_1 = N_2 = 6 \) to \( N_1 = N_2 = 20 \) (adding sequentially one new train from each direction). Finally, the number of trains was set to 25 and then 30 from each direction. With 30 train sin each direction the scheduling problem involves approximately 850 DEs.

The schedules for the case \( N_1 = N_2 = 6 \) and \( N_1 = N_2 = 30 \) are shown in Figure 1 and 2 in the standard scheduling diagram used in railway industry, with time displayed on the horizontal axis, and the distance (from O to D) on the vertical axis, with each trace representing the position of individual trains traveling from O to D, and from D to O. The horizontal lines represent the locations of the M&P points at which passes and overtakes can take place. A broken line at such a point represents a stop time for a particular train.

![Figure 1. N1 = N2 = 6](image1)

![Figure 2. N1 = N2 = 30](image2)

![Figure 3. Efficiency index \( \eta \) in function of \( N_1 = N_2 \)](image3)

While almost no decrease is to be noticed in the values of \( \eta \), the total delay is a quadratic function of \( N_1 = N_2 \), and indicates that traffic is approaching the capacity limit of the line with the considered type of traffic as the number of trains form each side increases beyond when \( N_1 = N_2 = 30 \).
Considering $\eta$ in function of the number of trains, it is observed that $\eta$ consistently remains above 0.95, Figure 3. Its mean value from this data is $\eta_m = 0.9861$ with a standard deviation of $\sigma = 0.0074$ (0.75% of the mean value). The results attest to the significant packing ability of the greedy TAS, as reflected in the value of $J_1$ and the efficiency ratio $\eta$.

5. DOUBLE TRACK SECTIONS

If some sections of the line have double tracks, the TAS is easily adapted by removing restrictions for M&P, and M&O in such sections. The scheduling problem, thus, has fewer conditions that define obstructions. The application of the greedy TAS then includes the list of double track section and the algorithm is appropriately modified. As illustrated by the example below the greedy TAS maintains its time-efficient nature with respect to $J_1$ and $\eta$.

Example 2. Consider the hypothetical issue of equipping one of the 11 sections in the case study considered in Example 1 into a double track section. Let $qq$ denote the section of the line with double tracks, with $qq = 0$ denoting the case when all sections have single tracks. For various $N_1 = N_2$ cases, a second track was introduced sequentially into each section of the line. The greedy strategy was then applied to obtain a schedule, and the effect noted on the relevant performance criteria. The analysis was repeated with shifted departure times of all trains traveling from D to O (using 0.5 – 2 h shifts in half hour increments) to avoid the possible biasing effects of a particular sets of departure times.

Shown in Table 2 is a sample analysis for $N_1 = N_2 = 30$, which tests the capacity of the line. The analysis was used to single out section of the line where addition of a second track would be most beneficial. All three time-efficiency related criteria defined earlier were considered. The analysis singled out the longest sections, corresponding to $qq = 6$ (30 mi), $qq = 10$ (25 mi), and $qq = 1$ (20 mi), in most cases. The introduction of a second track in any section usually, but not always, resulted in improved performance. The reason is that the local nature of the greedy TAS tends to exploit the second track although this may in some instances lead to worse performance in the remaining sections which cannot handle the increased volume of traffic. This opens new questions in terms of the performance of a local strategy with fewer (local) obstructions and these will be studied separately. Nevertheless, the performance with respect to the efficiency ratio $\eta$ remains excellent.

The study also provides evidence that the three time-efficiency related criteria used here are mutually independent, in that a decrease in value of one does not necessarily imply the decrease of the other two.

<table>
<thead>
<tr>
<th>$qq$</th>
<th>$\eta$</th>
<th>$J_1$</th>
<th>$J_2$</th>
<th>$J_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.907</td>
<td>27.9369</td>
<td>43.7412</td>
<td>1.7649</td>
</tr>
<tr>
<td>1</td>
<td>0.9984</td>
<td>27.7204</td>
<td>37.4129</td>
<td>1.7482</td>
</tr>
<tr>
<td>2</td>
<td>0.9805</td>
<td>28.2272</td>
<td>38.6141</td>
<td>1.3110</td>
</tr>
<tr>
<td>3</td>
<td>0.9910</td>
<td>27.9281</td>
<td>38.9677</td>
<td>1.2697</td>
</tr>
<tr>
<td>4</td>
<td>0.9883</td>
<td>28.0047</td>
<td>46.4320</td>
<td>1.8995</td>
</tr>
<tr>
<td>5</td>
<td>0.9910</td>
<td>27.9269</td>
<td>38.9508</td>
<td>1.6409</td>
</tr>
<tr>
<td>6</td>
<td>0.9949</td>
<td>27.8172</td>
<td>35.0030</td>
<td>1.8281</td>
</tr>
<tr>
<td>7</td>
<td>0.9823</td>
<td>28.1761</td>
<td>38.4445</td>
<td>1.2622</td>
</tr>
<tr>
<td>8</td>
<td>0.9886</td>
<td>27.9957</td>
<td>39.5246</td>
<td>1.8645</td>
</tr>
<tr>
<td>9</td>
<td>0.9880</td>
<td>28.0112</td>
<td>42.8741</td>
<td>1.7411</td>
</tr>
<tr>
<td>10</td>
<td>0.9892</td>
<td>27.9785</td>
<td>33.4953</td>
<td>1.2294</td>
</tr>
<tr>
<td>11</td>
<td>0.9907</td>
<td>27.9369</td>
<td>43.5693</td>
<td>1.7435</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

Local, state dependent, TAS and the discrete event models of a railway line represent a more efficient way of approaching the scheduling problem then nonlinear programming approaches. The onset of deadlock do to excessive traffic is easily recognizable because $d_{\text{next}}$ is reduced to zero. But, there simply is no line where a much greater number of trains from each direction needs to be scheduled. That is why local strategies work well. If deadlock is due to local conditions, it is easily avoided by shifting certain departure times, if it is not local, then modifying the TAS to include a larger local neighborhood can resolve it. This is currently being implemented in extending the TAS approach to railway networks.

REFERENCES


