Dynamic railway network management using switching max-plus-linear models

T.J.J. van den Boom and B. De Schutter

If you want to cite this report, please use the following reference instead:
Abstract

In this paper we discuss dynamic traffic management of railway networks. The main aim of the controller is to recover from delays in an optimal way by breaking connections and changing the departure of trains (at a cost). To model the railway system we use a switching max-plus-linear system description. We define the optimal control design problem for the railway network, and we show that solving this problem leads to an integer optimization problem. This problem can be solved with a genetic algorithm or with a mixed integer linear programming algorithm. We also apply the algorithm to a model of the Dutch railway network.

Keywords: Railway network management, switching max-plus-linear models, model predictive control

1. INTRODUCTION

In recent years a lot of research effort has been oriented towards the design of timetables that are robust against propagation of delays in the network, caused by technical failures, fluctuation of passenger volumes, measures of railway personnel and weather influence (Subiono, 2000; Hansen, 2001; Peeters, 2003; Goverde, 2005). In this paper we concentrate on the operational-level management, and design a feedback controller that takes the most effective actions, based on measurements of the actual train positions. The measures we can take are changing the train speed, breaking train connections, or changing the order of trains.

From (Braker, 1991; Braker, 1993b; de Vries et al., 1998; de Waal et al., 1997; Minciardi et al., 1995) we know that a railway network with rigid connection constraints and a fixed routing schedule can be modeled using max-plus-linear models. A max-plus-linear model is "linear" in the max-plus algebra (Baccelli et al., 1992), which has maximization and addition as its basic operations. Max-plus-linear systems can be characterized as discrete event systems in which only synchronization and no concurrency or choice occurs.

Note that in the railway context, synchronization means that some trains should give pre-defined connections to other trains, and a fixed routing means that the order of departure is fixed. However, in the case of large delays, it is sometimes better — from a global performance viewpoint — to break a connection or to reschedule the order of trains, and to let a train depart anyway. In this way we prevent an accumulation of delays in the network. Of course, missed connections should lead to a penalty due to dissatisfied passengers. In (De Schutter and van den Boom, 2001; De Schutter et al., 2002a) we have considered the control of railway networks using breaking connections only as control measure. In (van den Boom and De Schutter, 2004) we have extended the control handles

1 Corresponding author
and rescheduled the trains by breaking connections as well as changing train order. In this paper we will model a controlled railway system using the switching max-plus-linear system description of (van den Boom and De Schutter, 2006). In this description we use a number of MPL models, each model corresponds to a specific mode, describing the network by a different set of connection and order constraints. We control the system by switching between different modes, allowing us to break train connections and to change the order of trains. In this paper we define a control algorithm to optimize the performance of the network, and we show that the resulting optimization problem can be solved as a mixed integer problem or a mixed integer linear programming problem. Although these problems are in general NP-hard, recently several efficient solvers have become available. The management algorithm will be applied to a simulation model of the Dutch railway network. Computational experiments show that the proposed genetic algorithm approach yields good results.

2. MODEL

Consider a railway operations system, which follows a schedule with period $T$. In nominal operation mode, we assume that all the trains follow a pre-scheduled route, with fixed train order and pre-defined connections. If for some reason we have to break connections or change the train order, we will operate in a perturbed mode. With every new schedule we can associate a perturbed mode. First we will discuss the nominal operation.

2.1 Nominal operation

Consider a railway operations system which is operating in nominal operation mode.

Each track of the railway network has a number and a train allocated to it. For the sake of simplicity we will say ‘(virtual) train $j$’ to denote the (physical) train on track $j$, and ‘station $j$’ to denote the station at the beginning of track $j$ (cf. Figure 1). Let $n$ be the number of ‘virtual’ tracks in the network. We say virtual to denote that some of the virtual tracks may actually be the same physical track (corresponding to different trains using the same track). This means that the number of trains is usually smaller than $n$. Let $x_j(k)$ be the time instant at which train $j$ departs from station $j$ for the $k$th time. Let $d_j(k)$ be the departure time for this train according to the time schedule. Let $p_i(k)$ be the predecessor track of train $i$, and let $C_i(k)$ be the set of tracks to which the $i$th train $i$ gives a connection. Let $F_j(k)$ be the set of trains that move over the same track as train $j$, in the same direction as train $j$, and are scheduled behind train $j$. Let $V_j(k)$ be the set of trains that move over the same track as train $j$, in the opposite direction of train $j$, and are scheduled behind train $j$. Furthermore, let $a_j(k)$ be the traveling time on track $j$, define a minimum connection time $c_{ij}^{\text{min}}(k)$ for passengers to get from train $j$ to train $i$ for each train $j \in C_i(k)$ and define a minimum stopping time $s_j^{\text{min}}(k)$ of train $j$ at station $j$ to allow passengers to get off or on the train. Finally, define a minimum separation time $f_{ij}^{\text{min}}(k)$ between two different trains moving over the same track and in the same direction as train $j$, and a minimum separation time $w_{ij}^{\text{min}}(k)$ between two different trains moving over the same track and in the opposite direction.

Now we have the following constraints for the $k$th departure time $x_i(k)$ of train $i$:

- **Time schedule constraint:**
  
  \[ x_i(k) \geq d_i(k) \, . \]

- **Continuity constraints:** This constraint synchronizes two trains that are ‘physically’ the same train. For train $j = p_i(k)$ we have
  
  \[ x_i(k) \geq x_j(k) - \delta_{ij}^s(k) + a_j(k) + s_j^{\text{min}}(k) \]

  where $\delta_{ij}^s(k)$ is equal to 1 if the $(k - 1)$th train $j$ continues as the $k$th train $i$, and 0 if the $k$th train $j$ continues as the $k$th train $i$ (and if some trips last longer than twice the cycle time $T$ of the schedule, $\delta_{ij}^s(k)$ might be equal to 2, and so on — see also the example in Section 4). In general, $\delta_{ij}^s(k)$ may depend on $k$. However, for the sake of simplicity, we only consider constant $\delta_{ij}^s(k)$’s with a value that is either 0 or 1 in this paper.

- **Connection constraints:** This constraint synchronizes two trains that have to make a connection. For each train $i \in C_j(k)$ we have
  
  \[ x_i(k) \geq x_j(k) - \delta_{ij}^s(k) + a_j(k) + c_{ij}^{\text{min}}(k) \]

  where the role of $\delta_{ij}^s(k)$ is similar as for the continuity constraint, so $\delta_{ij}^s(k) = 1$ if the $(k - 1)$th train $j$ gives a connection to the $k$th train $i$, and $\delta_{ij}^s(k) = 0$ if the $k$th train $j$ gives a connection to the $k$th train $i$.

- **Follow constraints:** This constraint synchronizes two subsequent trains on the same track moving in the same direction. For each train $i \in F_j(k)$ we have
  
  \[ x_i(k) \geq x_j(k) - \delta_{ij}^s(k) + f_{ij}^{\text{min}}(k) \]

  ($\delta_{ij}^s(k)$ is defined similarly as above).

- **Wait constraints:** This constraint synchronizes two trains on the same track moving in opposite direction. For each train $i \in V_j(k)$ we have
  
  \[ x_i(k) \geq x_j(k) - \delta_{ij}^s(k) + a_j(k) + w_{ij}^{\text{min}}(k) \]

Figure 1. A part of a railway network.
(\delta_{ij}^* \text{ is defined similarly as above}).

Since we let a train depart as soon as all connection conditions are satisfied, we have
\[ x_i(k) = \max(d_i(k), \max_{j \in C_i}(x_{ij}(k-\delta_{ij}^*) + a_{ij}(k) + c_{ij}^{\min}(k)), \]
\[ \max_{j \in C_i}(x_{ij}(k-\delta_{ij}^*) + a_{ij}(k) + c_{ij}^{\min}(k)), \]
\[ \max_{k \in T_i}(x_k(k-\delta_{ik}^*) + f_{ik}^{\min}(k)), \]
\[ \max_{m \in W_i(k)}(x_m(k-\delta_{im}^*) + a_m(k) + w_{im}^{\min}(k))) \]
\[ x_i(k) = \max(d_i(k), \max_{j \in C_i}(x_{ij}(k-m) + [A^{0}_m]_{i,j})) \]
\[ (2) \]

Note that in a undisturbed, well-defined time schedule the term \(d_i(k)\) in (1) will be the largest. However, if due to unforeseen circumstances (an accident, a late departure, etc.) one of the trains \(p_i(k), I \text{ or } m\) has a delay the corresponding term can become larger than the others, then train \(i\) will depart later than the scheduled departure time \(d_i(k)\) and will therefore also be delayed. By defining the appropriate matrix \(A_m\), \(m = 0, \ldots, m_{\text{max}}\) (where \(m_{\text{max}} = \max(\delta_{ij}^*)\) we can rewrite equation (1) as:
\[ x_i(k) = \max\left(d_i(k), \max_{j,m}(x_{ij}(k-m) + [A^0_m]_{i,j})\right) \]
\[ \text{where } [A^0_m]_{i,j} \text{ is the } (i,j) \text{th entry of the matrix } A^0_m. \]

Now we introduce some notation from max-plus algebra. Define \(\varepsilon = -\infty\) and \(\mathbb{R}_\varepsilon = \mathbb{R} \cup \{\varepsilon\}\). The max-plus-algebraic addition \((\oplus)\) and multiplication \((\otimes)\) are defined as follows (Baccelli et al., 1992):
\[ x \oplus y = \max(x, y) \quad x \otimes y = x + y \]
for \(x, y \in \mathbb{R}_\varepsilon\) and
\[ [A \oplus B]_{i,j} = a_{ij} \oplus b_{ij} = \max(a_{ij}, b_{ij}) \]
\[ [A \otimes C]_{i,j} = \bigoplus_{k=1}^{n} a_{ik} \otimes c_{kj} = \max_{k=1}^{n} (a_{ik} + c_{kj}) \]
for \(A, B \in \mathbb{R}^{m \times n}, C \in \mathbb{R}^{n \times p}\). The matrix \(E\) is the max-plus-algebraic zero matrix: \([E]_{i,j} = \varepsilon\) for all \(i,j\).

In max-plus notation, equation (2) becomes
\[ x_i(k) = d_i(k) \oplus \bigoplus_{j=1}^{n} \bigotimes_{m=1}^{M} x_{ij}(k-m) \otimes [A^0_m]_{i,j} \]
and in matrix notation we obtain
\[ x(k) = A^0_0 \otimes x(k) + A^0_1 \otimes x(k-1) \oplus \ldots \]
\[ \oplus [A^0]_{m_{\text{max}}} \otimes x(k-m_{\text{max}}) \oplus d(k) \]
\[ = \left( \bigoplus_{m=0}^{m_{\text{max}}} A^0_m \otimes x(k-m) \right) \oplus d(k) \]
\[ (3) \]

2.2 Perturbed operation

In the nominal operation we have assumed that some trains should give pre-defined connections to other trains, and the order of trains on the same track is fixed. However, if one of the preceding trains has a too large delay, then it is sometimes better — from a global performance viewpoint — to let a connecting train depart anyway or to change the departure order on a specific track. This is done in order to prevent an accumulation of delays in the network. In this paper we will consider the switching between different operation modes, where each mode corresponds to a different set of pre-defined or broken connection and a specific order of train departures. We allow the system to switch between different modes, allowing us to break train connections and to change the order of trains. Note that any broken connection or change of train order leads to a new model, similar to the nominal equation (3), but now with adapted system matrices \((A^\ell)\) for the \(\ell\)-th model. We have the following system equation for the perturbed operation\(^2\):
\[ x(k) = \left( \bigoplus_{m=m_{\text{min}}}^{m_{\text{max}}} A^\ell_m \otimes x(k-m) \right) \oplus d(k) \]
\[ (4) \]

3. THE RAILWAY CONTROL PROBLEM

3.1 Timing aspects

Switching max-plus-linear systems are different from conventional time-driven systems in the sense that the event counter \(k\) is not directly related to a specific time. A time instant \(t\) in cycle \(k\) (so \((k-1)T \leq t < kT\)), some of the components of \(x(k)\) may already be known while other components of \(x(k)\) may still lie in the future (Recall that \(x(k)\) contains the time instants at which the internal activities or processes of the system start for the \(k\)-th cycle). Therefore, we will now present a method to address the timing issues in control of switching MPL systems.

We consider the case of full state information\(^3\), since the components of \(x(k)\) correspond to departure times, which are in general easy to measure.

Consider time instant \(t\) in cycle \(k\), so \((k-1)T \leq t < kT\). We have measurements of departure times \(x_{\text{past}}(k)\) and traveling times \(\delta_{\text{past}}(k)\) of trains that have arrived at their destination. Sometimes there is information about the estimated future traveling time for trains that have not yet arrived at their destination. If no further information is available on a specific traveling time we take the nominal traveling time \([\hat{\alpha}_{\text{est}}(k|l)]_i = \alpha_{i,\text{nom}}\).

\(^2\) Usually \(m_{\text{min}} = 0\). However, in perturbed operation it may occasionally happen that a delayed train of the \(k\)-th cycle is rescheduled behind a train in the \((k+1)\)-th cycle. In that case we will have \(m_{\text{min}} = -1\).

\(^3\) Note that measurements of occurrence times of events are in general not as susceptible to noise and measurement errors as measurements of continuous-time signals involving variables such as temperature, speed, pressure, etc.
3.2 Control problem

Next we have to define the set $\mathcal{U}(k|t)$ of possible future control actions (i.e. breaking connections or changing train order). Certain control actions are not feasible anymore (e.g. if a connection has been broken in the past and the connecting train has already departed, it is impossible to ‘repair’ this connection.). We define the vector $u(k|t) \in \mathcal{U}(k|t)$, where each element corresponds to a specific control action, so a specific (scheduled) connection or specific (scheduled) train order. We assume $u(k|t)$ to be binary, where $u_i(k|t) = 0$ corresponds to the nominal case, and $u_i(k|t) = 1$ to the perturbed case (the connection is broken or the order of two trains is switched, see also Section 3.3).

To select the optimal set of possible future control actions, we define the following optimal control problem at time instant $t$ ($((k-1)T \leq t < kT)$:

$$\min \{ u(k|t), u(k+1|t), u(k+2|t), \ldots \} \ J(k|t)$$

(5)

where the performance index $J(k|t)$ is given by

$$J(k,t) = \sum_{j=0}^{\infty} \sum_{i=1}^{n} Q_i \hat{e}_i(k+j|t) + \sum_{\ell=1}^{n_u} R_{\ell} u_{\ell}(k+j|t)$$

(6)

where $\hat{e}(k+j|t)$ is the vector with the expected delays ($\hat{e}_i(k+j|t) = \hat{x}_j(k+j|t) - d_i(k+j) \geq 0$, and $Q, R$ are weighting matrices. The first term of (6) is related to the sum of all predicted delays, and the second term denotes the penalty for all broken connections and switched train orders during cycle $k + j$.

To compute the predictions of $\hat{x}(k+j|t)$ we make use of the fact that at time $t$ we have $a_{\text{dat}}(k|t)$ and $\hat{a}_{\text{dat}}(k+j|t)$ available and using that we can determine the estimates $\hat{A}(k+j|t)$ of all future $A(k+j)$. Now $\hat{x}(k+j|t)$ for $j \geq 1$ can be found by successive substitution

$$\hat{x}(k+j|t) = \hat{A}(k+j-1|t) \odot \hat{x}(k+j-1|t) + d(k+j)$$

(7)

In principle we have all elements to solve the optimal control problem (5). Note that if the railway network is well-defined and there is some margin in the schedule, there is ‘some margin’ in the schedule. We have always an integer $N$ such that in the nominal case ($u(k+j|t) = 0$ for all $j \geq 0$) the delays will have vanished ($\hat{e}(k+j|t) = 0$ for all $j \geq 0$). In the performance index (6) we may then replace the infinite sum by a finite one (with an optional constraint $\hat{e}(k+N|t) = 0$). We now have an integer optimal control problem with $nN$ binary parameters. We can solve this problem efficiently with genetic algorithms (Davis, 1991) or with tabu search (Glover and Laguna, 1997).

3.3 Initial solution

To find a good initial guess for the integer optimization we first solve an easier problem, in which we structure the input signal. This is done by defining a decision mechanism, where we use thresholds on (expected) delays to decide whether a connection should be broken or train orders should be switched. First consider the case where variable $u_i(k)$ is related to the connection of train $j$ to train $i$, with nominal connection constraint

$$x_i(k) \geq x_j(k - \delta_{ij}^*) + a_j(k) + c_{ij}^\text{min}(k)$$

and let $d_i(k) > t$. Define $\hat{z}_j(k - \delta_{ij}^*)|t| = \hat{x}_j(k - \delta_{ij}^*)|t| + [a_{\text{est}}]_j(k|t)$ as the expected arrival-time of train $j$. Now we choose

$$\{ u_i(k) = 0 \quad \text{if} \quad \hat{z}_j(k - \delta_{ij}^*)|t| + c_{ij}^\text{min}(k) - d_i(k) \leq \tau \}
$$

$$u_i(k) = 1 \quad \text{otherwise,}$$

where $\tau$ is a non-negative threshold. Next consider the case where variable $u_i(k)$ is related to the order of two trains $j$ and $i$ moving over the same track in the same direction, with nominal following constraint

$$x_i(k) \geq x_j(k - \delta_{ij}^*) + f_{ij}^\text{min}(k)$$

and let $x_j(k) \geq t$ (that means that at time $t$ train $x_i(k)$ has not departed yet). Now we choose

$$\{ u_i(k) = 0 \quad \text{if} \quad \hat{z}_j(k - \delta_{ij}^*)|t| + f_{ij}^\text{min}(k) - d_i(k) \leq \phi \}
$$

$$u_i(k) = 1 \quad \text{otherwise,}$$

where $\phi$ is a non-negative threshold. Finally consider the case where variable $u_i(k)$ is related to the order of two trains $j$ and $i$ moving over the same track in the same direction, with nominal waiting constraint

$$x_i(k) \geq x_j(k - \delta_{ij}^*) + a_j(k)$$

and let $d_i(k) > t$. Now we choose

$$\{ u_i(k) = 0 \quad \text{if} \quad \hat{z}_j(k - \delta_{ij}^*)|t| - d_i(k) \leq \omega \}
$$

$$u_i(k) = 1 \quad \text{otherwise,}$$

where $\hat{z}_j(k - \delta_{ij}^*)|t|$ is the expected arrival-time and $\omega$ is a non-negative threshold. In this structured-input case we end up with the minimization of (6) using the three parameters, giving us a non-linear optimization problem over the variables ($\tau, \phi, \omega$). In the worked example in the next Section we first optimize over the structured inputs, and use the resulting sequence $u(k+j|t)$ as an initial value for the general case, solved with a genetic algorithm.

Remark: The problem above can also be recast as a mixed integer linear programming problem (MILP) using techniques that are similar to the ones used in (Bemporad and Morari, 1999; De Schutter et al., 2002b). We will now briefly outline the main ideas behind this transformation. Note that the objective function $J$ is linear in $u$ and $\hat{x}$ (via $\hat{e}$). The max-plus equation (7) can be transformed into a system of mixed-integer linear inequalities as follows. Consider an equation of the form $\alpha = \max(\beta, \gamma)$. Hence, we have $\alpha - \beta \geq 0$ and $\alpha - \gamma \geq 0$. Now assume that

---

4 If the max-plus eigenvalue (Braker, 1993a) of the matrix $A^0$ is strictly smaller than $T$, there is ‘some margin’ in the schedule.
We consider a control design method for a railway system. We first optimize the threshold values \( u \) for the optimal controlled case, and then we compute the corresponding optimal structured input signal \( u_{\text{structured}}(k+j|t) \), \( j \geq 0 \). Subsequently we optimize the (unstructured) input signal \( u(k+j|t) \) with a genetic algorithm, using the earlier computed sequence \( u_{\text{structured}}(k+j|t) \) as an initial value.

In Figure 3 the maximum delay \( e_{\text{max}}(k) = \max(e(k)) \) in each cycle \( k \) is given for the optimal controlled case, and for the uncontrolled case (so \( u(k+j|t) = 0 \) for all \( j > 0 \)). We see that the delay in the controlled case decays much faster than the uncontrolled case.

5. DISCUSSION

We have presented a control design method for a railway network. The control action consists in breaking certain connections or changing the order of departure to prevent delays from accumulating. These control moves can only be done at a certain cost. We have shown that the resulting optimization problem can be
Figure 3. Maximum delay for uncontrolled and optimal controlled railway system

solved using integer optimization methods, for example genetic algorithms or tabu search.

Good initial values for the integer optimization are obtained by first solving a low-dimensional real-valued optimization problem using a structured input sequence. This structured input sequence is based on a decision mechanism, where we use thresholds on (expected) delays to decide whether a connection should be broken or the order of the trains should be switched.

Due to the use of a receding horizon this method can be used in on-line applications and it can deal with (predicted) changes in the system parameters. So if we can predict the delays that will occur due to an incident or to works, then we can include this information when determining the optimal control input for the next cycles of the operation of the network.

Acknowledgments

Research partially funded by the Dutch Technology Foundation STW project ‘Model predictive control for hybrid systems’ (DMR.5675), the Transport Research Centre Delft, the European Network of Excellence HYCON (FP6-IST-511368), and by the NWO/STW VIDI project ‘Multi-agent control of large-scale hybrid systems’ (DWV.6188).

REFERENCES


