Transactions on the Built Environment vol 34, © 1998 WIT Press, www.witpress.com, ISSN 1743-3509

The max-plus algebra approach to railway timetable design

R.M.P. Goverde
Faculty of Civil Engineering and Geo Sciences, Delft University of Technology, PO Box 5048, 2600 GA Delft, The Netherlands
email: goverde@ct.tudelft.nl

Abstract

A scheduled railway system that operates according to a cyclic timetable naturally exhibits a cyclic (periodic) behaviour. In a max-plus algebra setting such a system can be modelled as a linear (discrete event) dynamic system. The computation of a timetable then reduces to solving an eigenvalue problem for which efficient algorithms have been developed. Moreover, the max-plus algebra system theory contains stability analysis and simulation facilities. This paper shows that the max-plus algebra approach offers an efficient interactive timetable design framework which directs attention to the critical components in the railway system.

1 Introduction

An operational timetable must be robust to uncertainty in train running times and corresponding arrival delays. A too tight schedule results in missed transfers and/or perturbations in train movements. Connection buffer times between arrival and departure times of connected trains reduce arrival delays and increase transfer reliability. Both from an operational and passenger point of view. However, again from both perspectives connection times should not be too large since this results in an increase of travel times and rolling stock demand, and a reduction of capacity utilization. The determination of optimal connection times solving this dilemma is thus of major concern for all parties involved: passengers, process operators, and train service providers.

Mathematical programming offers an obvious approach for solving timetable design problems. The problem is then to find optimal feasible arrival and departure times satisfying a set of constraints such that an objective function (e.g., the sum of expected transfer waiting times) is minimized. The constraints correspond to lower and/or upper bounds on, e.g., running times, connection times and headways. The resulting problems are mixed integer program (MIP) problems, see for instance [2, 6, 9, 10, 11]. However, solving these various problem formulations is computationally very hard for large networks. In fact these MIP problems are all NP-complete. This is inherently due to the periodicity of the cyclic timetable which results in objective functions and/or constraints that are relative to modular arithmetic (addition and subtraction are modulo the cycle time). For large complex train service networks as in the Netherlands this leads to intractable problems.

Dynamic system theory offers an alternative approach for modelling railway systems. The periodicity is here implicitly contained in the dynamic equation describing the interactions between the various train movements. The dynamic behaviour of railway systems is naturally event driven. Events (e.g., train arrivals and departures) are subject to precedence constraints. Such systems are discrete event dynamic systems (DEDS) which can be modelled using the max-plus algebra [3, 7]. Stability and robustness can efficiently be analysed in max-plus algebra systems [3]. Moreover, a max-plus algebra model can be used to simulate delay propagation in the service network.

Recently, it has been shown that computing a timetable equals solving an eigenvalue problem in the max-plus algebra [3]. Within the max-plus algebra framework a timetable designer has thus the opportunity to compute candidate timetables, analyse their stability, and simulate delay propagation to test sensitivity/robustness to delays.

This paper shows how train service networks can be modelled as max-plus algebra systems and describes an efficient approach to design timetables.

2 The Precedence Graph

The main problem in the design of timetables for a train service network is the scheduling of connections between individual trains at transfer stations. This paper therefore assumes that the running times between transfer stations, including the stopping times at intermediate stops, are known and concentrates on the connections.

Of interest is not the physical railway network but the connection network or *precedence graph.* This is a network representation of precedence constraints.

The precedence graph is a directed graph consisting of a set of n nodes and a set of m arcs. The nodes corresponds to trains or train departures at transfer stations, and the arcs represent precedence constraints corresponding to connections between trains at transfer stations. A connection is either a physical connection or a transfer connection. A weight a_{ji} is assigned to each arc (i, j) corresponding to the sum of the running time t_i^r of train i and the connection time from train i to j. Here, the connection time is either a stopping time t_{ij}^s or a transfer time (or changeover time) t_{ij}^c. The fact that a weight a_{ij} corresponds to an arc (j, i) may be confusing at first sight. However, in Section 4 it will be shown that by doing so, the railway system can be formulated as a familiar linear system $x(k+1) = Ax(k)$.

Figure 1 shows an example train service network consisting of two transfer stations S_1 and S_2 and 4 routes (the routes are indicated in bold numbers). Intermediate stops along the routes have not been drawn. The system has three lines (train series): a line serving route 1, a line connecting the transfer stations in both directions (serving routes 2 and 3), and a line serving route 4. The weights at the arcs indicate the running times of the routes, and the weights around the nodes (transfer stations) indicate the minimum stopping or transfer times between the arriving and departing arcs. So, trains circulating on the routes 2 and 3 have a minimum stopping time of 1 minute at the transfer stations, and trains of route 1 and 4 have a stopping time of 3 minutes at the transfer stations. The minimum transfer times are all 2 minutes. Figure 2 shows the corresponding precedence graph.

We assume that the graph is strongly connected, i.e., there is a (directed) path between any node i to any node j, where a path is a sequence of adjacent nodes (without any repetition of nodes). If this is not the case then the graph can be partitioned into strongly connected subgraphs which are treated separately. Note that a strongly connected graph inherently contains circuits, where a *circuit* is a (closed) path where the end points coincide. The weight of a circuit is the sum of its arc weights, i.e., the sum of running and connection times. As an example, the graph of Figure 2 contains 6 circuits. The circuit of the successive nodes 1-3-4-2-1 has weight $52+43+28+44 = 167$. The occurrence of circuits makes cyclic timetable design nontrivial: the circuit weights must equal an integer multiple of the cycle time [11].

The *cycle mean* of a circuit is the average trip time on the circuit,

Transactions on the Built Environment vol 34, © 1998 WIT Press, www.witpress.com, ISSN 1743-3509

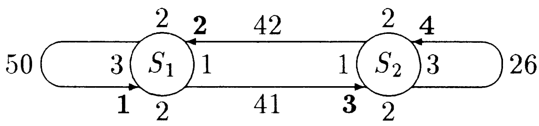

Figure 1: The train service network

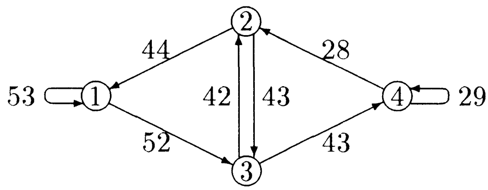

Figure 2: The precedence graph

i.e., the ratio of circuit weight and length, where the *circuit length* is the number of arcs (trains) in the circuit. For example, the cycle mean of circuit 1-3-4-2-1 has cycle mean $167/4 = 41\frac{3}{4}$. A *critical circuit* is a circuit with *maximum cycle mean* (where the maximum is taken over all circuits). The maximum cycle mean of the graph in Figure 2 is 53 corresponding to circuit 1-1.

Note that the minimal interdeparture time of trains on the critical circuit equals the circuit's (maximum) cycle mean for suitable chosen departure times. The critical circuit is thus the 'slowest' circuit in the train service network. This implies the following important observation. The maximum cycle mean is the *critical cycle time*, or minimum cycle time, of the (overall) timetable.

3 Discrete Event Railway Systems

A discrete event dynamic system description of a railway system is easily obtained from the precedence graph.

Consider the precedence graph. Assign to each node i a departure time x_i. This departure time depends on the arrival time from the train's preceding trip as well as on the arrival times of its feeder trains. The earliest possible departure time of a train i is therefore formally given as

$$x_i = \max_{j=1,\dots,n} (a_{ij} + x_j), \qquad i = 1, \dots, n, \tag{1}$$

where a_{ij} is defined as the sum of the running time t^r_j, and the stop-

Transactions on the Built Environment vol 34, © 1998 WIT Press, www.witpress.com, ISSN 1743-3509

ping time t^s_{ji} or transfer time t^c_{ji} between trips j and i, respectively,

$$a_{ij} = \begin{cases} t^r_j + t^s_{ji} & \text{if train } j \text{ physically equals train } i \\ t^r_j + t^c_{ji} & \text{if train } j \text{ is a feeder train of train } i \\ -\infty & \text{otherwise.} \end{cases} \tag{2}$$

Note that assigning $-\infty$ to trains that are not connected to train i implies that these trains can also be incorporated in eqn (1) since these entries have no influence on the maximization (as long as other trains have finite entries). The trains j for which $a_{ij} \neq -\infty$ correspond to the predecessors of node i in the precedence graph.

We are concerned with a cyclic timetable. Therefore the departure times are periodic recurrent events. Let k be a counter denoting a specific period. Then the kth departure time of a train series i is $x_i(k)$. Incorporating the periodicity in eqn (1) gives

$$x_i(k+1) = \max_{j=1,\ldots,n} \left(a_{ij} + x_j(k)\right), \qquad i = 1,\ldots,n. \tag{3}$$

The departure time of a train i thus depends on former departure times of the preceding trains. In general a train may also be connected to a train that departed two or more periods before. However, this situation can be reduced to eqn (3) by using an *augmented precedence graph*, see Section 6.

If the train service network operates according to a timetable then a train may not depart before its scheduled departure time. However, if the train is behind schedule, or has to wait for a delayed feeder train, then the actual departure time may exceed the scheduled departure time. Denote the scheduled departure time of a train i from a transfer station as d_i. Then the scheduled train service network can be described as

$$x_i(k+1) = \max\left(a_{i1} + x_1(k), \ldots, a_{in} + x_n(k), d_i(k+1)\right), \tag{4}$$

for $i = 1,\ldots,n$. The subsequent scheduled departure times for train i are given as

$$d_i(k) = d_i(0) + kT,$$

where T is the cycle time and $d_i(0)$ is an initial departure time.

A timetable naturally contains *connection buffer times* defined as the intervals between the earliest possible departure times and the scheduled departure times at transfer stations. From eqn (4) follows that the connection buffer time r_{ji} between an arriving train j and a departing train i is

$$r_{ji}(k+1) = d_i(k+1) - (a_{ij} + x_j(k)). \tag{5}$$

If initial departure times $x_1(0), \ldots, x_n(0)$ are given then the evolution of the railway system (3), or the scheduled railway system (4), is completely determined, i.e., the subsequent departure times of all trains are uniquely fixed. These systems are examples of Discrete Event Dynamic Systems (DEDS). Here, a discrete event is a departure at a transfer station that occurs at a discrete instance in time, the departure time, and the dynamic equation, eqn (3) or eqn (4), describes the dynamic behaviour over the successive periods k. The above described systems are deterministic. If the parameters a_{ij} also depend on the period k then the system is stochastic. The subsequent running and connection times are then variable.

4 Max-Plus Algebra

Eqn (3) contains the two operations maximization and addition which makes it nonlinear in a linear algebra sense. However, with a change of notation eqn (3) can be written as a linear system. For this, denote maximization as $\oplus$, and addition as $\otimes$. Thus, $x \oplus y = \max(x, y)$ and $x \otimes y = x + y$. Then eqn (3) becomes

$$x_i(k+1) = \bigoplus_{j=1}^{n} (a_{ij} \otimes x_j(k)), \qquad i = 1, \ldots, n, \tag{6}$$

where $\oplus_{j=1}^{n} x_j = \max(x_1, \ldots, x_n)$ denotes repeated maximization. In vector notation eqn (6) is written as

$$x(k+1) = A \otimes x(k), \tag{7}$$

where $x = (x_1, \ldots, x_n)'$ and A is the square $n \times n$ matrix whose ijth entry is a_{ij}. Note the resemblance with linear algebra where a matrix equation $b = Ax$ is defined as $b_i = \sum_{j=1}^{n} (a_{ij} \cdot x_j)$ for $i = 1, \ldots, n$. Of course, this is no coincidence. Eqn (7) is a linear system in the *max-plus algebra*.

The max-plus algebra is like the conventional linear algebra but, as introduced above, the addition is replaced by maximization, denoted as $\oplus$, and multiplication is replaced by the conventional addition, denoted as $\otimes$. The set of elements considered in the max-plus algebra are the real numbers $\mathbb{R}$ and the additional element $\epsilon := -\infty$.

Note that the elements a_{ij} defined in eqn (2) belong to this set. The extension to vectors is equivalent to the linear algebra: the addition of two vectors is defined componentwise and matrix equations are defined analog to eqn (6). Concepts from linear algebra and linear system theory have their counterparts in the max-plus algebra [1, 5]. This paper does not give an extensive treatment of the max-plus algebra but restricts to illustrate its potential in the application to railway timetable design. The max-plus algebra modelling and (stability) analysis of (scheduled) railway systems is due to Braker [3], see also Goverde *et al.* [7].

The railway system (3) is thus a linear system in the max-plus algebra which in matrix notation is $x(k+1) = A \otimes x(k)$ or simply $x(k+1) = Ax(k)$. The matrix A is called the *state matrix* and the vector x is the state vector. For example, the matrix A corresponding to the precedence graph of Figure 2 is

$$A = \begin{pmatrix} 53 & 44 & \epsilon & \epsilon \\ \epsilon & \epsilon & 42 & 28 \\ 52 & 43 & \epsilon & \epsilon \\ \epsilon & \epsilon & 43 & 29 \end{pmatrix}. \tag{8}$$

Also the scheduled railway system (4) is a linear system in the max-plus algebra,

$$x(k+1) = Ax(k) \oplus d(k+1), \tag{9}$$

where $d = (d_1, \ldots, d_n)'$.

5 Critical Cycle Time and Timetable

The *critical cycle time* of railway system (7) is the minimum cycle time for which a realizable timetable exists. The corresponding timetable is a vector of scheduled departure times for all trains. The main advantage of the max-plus algebra modelling is the following result: the critical cycle time and timetable of the railway system are equivalent to the eigenvalue and eigenvector of the A matrix of eqn (7), respectively.

Let A be a square matrix in the max-plus algebra. Consider the eigenvalue problem: find a scalar λ and a vector $v = (v_1, \ldots, v_n)' \neq (\epsilon, \ldots, \epsilon)'$ such that

$$A \otimes v = \lambda \otimes v. \tag{10}$$

If this equation has a solution then λ is called the eigenvalue and v

an eigenvector. In conventional algebra eqn (10) becomes

$$\max_{j=1,\dots,n} (a_{ij} + v_j) = \lambda + v_i, \qquad i = 1, \dots, n. \tag{11}$$

This equation can be interpreted as follows. Assume that v is the vector of departure times in the present period. Then the earliest possible departure times in the next period (the left-hand side of eqn (11)) equals a constant λ added to the present departure times (the right-hand side of eqn (11)). The eigenvalue is thus the critical cycle time for which a cyclic timetable exists and the eigenvector v is a timetable for which this critical regular behaviour is satisfied. Note that the eigenvector is not unique: (conventional) addition of a constant to all components v_i also gives a vector satisfying eqn (11).

A main result in the max-plus algebra system theory is the following. If the precedence graph $G(A)$ is strongly connected then there exists a unique eigenvalue and at least one eigenvector, and the eigenvalue equals the maximum cycle mean of $G(A)$. Note that Section 2 already showed the interpretation of the maximum cycle mean as critical cycle time.

The example state matrix A has eigenvalue $\lambda = 53$ and eigenvector $v = (12, 0, 11, 1)'$ which can easily checked by eqn (11).

The eigenvalue, the eigenvector, and the corresponding critical circuit can be computed by an (extended) power algorithm [4, 3]. The computational complexity depends on the transient behaviour of the system [3, 5]. Karp's Algorithm [8, 3] efficiently computes the eigenvalue and critical circuit in $O(nm)$ time, where n and m are the number of nodes (trains) and arcs (connections) in the precedence graph, respectively.

6 A Timetable Design Approach

This section describes a timetable design approach for computing stable timetables for a minimum amount of rolling stock.

It is assumed that a line system is given. A line is a route between two terminal stations on which trains run with fixed frequency and serve given stops along the route. A line system consists of all lines characteristics and a set of connections between individual lines at transfer stations where the lines meet. Moreover, it is assumed that the running times between transfer stations (including stopping times at intermediate stops), and the minimum stopping times and minimum transfer times at the transfer stations are predetermined.

Transactions on the Built Environment vol 34, © 1998 WIT Press, www.witpress.com, ISSN 1743-3509

A timetable designer has to compute a timetable with a desired cycle time T, e.g., $T = 30$ minutes. A matrix A can be defined from the line system data by eqn (2) which results in the max-plus algebra railway model (7). Computing the eigenvalue of A gives the minimum possible cycle time λ. If $T > \lambda$ then this implies that with the present amount of trains it is not possible to operate according to a timetable with cycle time T. Therefore, an additional train has to be assigned to the system such that λ decreases.

Consider an edge (j, i) on the critical circuit in the train service network. Add a node l between node j and i and replace the arc (j, i) with two arcs (j, l) and (l, i), and divide the original arc weight a_{ij} over the new arcs. This node l can be interpreted as follows. In a particular period k, train j has a connection to train l at a fictitious station (with zero connection time), and the next period this train l has a connection to train i. In this way, train i has to wait for a train j with a departure of two periods before, i.e., the departure time $x_i(k+1)$ must be not earlier than $x_j(k-1) + a_{ij}$.

The assignment of an extra train to the critical circuit results in a smaller cycle mean of this circuit (recall that the cycle mean is the ratio of the total circulation time on the circuit and the number of circulating trains on the circuit). Now, for the matrix corresponding to the augmented precedence graph a new critical circuit can be computed with corresponding maximum cycle mean. The process of adding an extra train to the critical circuit of the subsequent augmented precedence graphs repeats until an eigenvalue is obtained for which $\lambda \leq T$. For the resulting railway system a realizable timetable exists with cycle time T. This approach results in a timetable that utilizes a minimum amount of rolling stock.

A timetable can be computed as the eigenvector of the augmented state matrix A (with corresponding eigenvalue λ). Using the cycle time T instead of λ results in connection buffer times (5) which implies that the timetable is stable: connection buffer times reduce delay propagation in the service network. The resulting scheduled railway system (9) can be analyzed on performance by means of stability analysis and/or simulation of the propagation of delays. If performance should be improved then extra trains can be assigned to those lines that are part of circuits in the train service network where delay reduction is not satisfactory.

The candidate timetable can still be tuned by adjusting the arrival times within the existing slack of running time margins and connection buffer times without consequences to the cycle time and

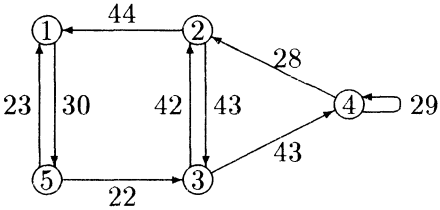

Figure 3: The augmented precedence graph, 1st iteration

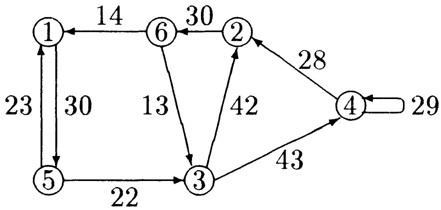

Figure 4: The augmented precedence graph, 2nd iteration

stability.

As an example, we will compute a timetable with cycle time $T = 30$ minutes for the railway system of Figure 1 with state matrix A given in eqn (8). The eigenvalue of A is $\lambda = 53$ corresponding to the critical circuit 1-1. We assign an extra train to the train series circulating on route 1. With respect to Figure 1, this corresponds to a fictitious station on route 1. Figure 3 shows the augmented precedence graph. The system now contains 5 trains. The new maximum cycle mean is $\lambda = 42\frac{1}{2}$ corresponding to circuit 2-3-2. Assigning an extra train to the train series circulating on routes 2 and 3 gives the augmented precedence graph of Figure 4. The railway system now contains 6 trains and the new maximum cycle mean is $\lambda = 29$ corresponding to circuit 4-4. Thus, a timetable can be computed with a minimum cycle time of 29 minutes with 6 circulating trains. A timetable can be computed as the eigenvector of the augmented state matrix corresponding to the precedence graph of Figure 4. This gives

$$d = (1, 15, 0, 16, 2, 16)'.$$

The first four components of this vector are the scheduled departure times of the train series at the transfer stations. The last two components are dummy departure times at the fictitious stations corresponding to the auxiliary trains. These are necessary for the max-plus algebra (simulation) model (9). Additionally to the state vector x of the departure times of all trains in the system (including auxiliary trains), an *output vector* y can be defined consisting of the departure times of the trains from (physical) transfer stations only. In this way, the user is not bothered with the auxiliary variables.

7 Conclusions and Future Research

A stable timetable can be computed efficiently using a power algorithm. Moreover, a critical (minimum) cycle time can be computed for which a realizable timetable exists with respect to the number of trains assigned to the lines (train series). Addition of trains to lines that are part of a critical circuit decreases the critical cycle time of the network. This can be used to design realizable timetables with a minimum amount of rolling stock and an optimal distribution of the rolling stock to lines with respect to sensitivity to delays.

Present research focusses on the computation of optimal connection buffer times within the max-plus algebra framework.

Acknowledgement

This publication is a result of the research programme Seamless Multimodal Mobility, carried out within TRAIL Research School for Transport, Infrastructure and Logistics, and financed by Delft University of Technology.

References

[1] Baccelli, F.L., Cohen, G., Olsder, G.J. & Quadrat, J.-P., *Synchronization and Linearity: an Algebra for Discrete Event Systems*, Wiley, Chicester, 1992.

[2] Bookbinder, J.H. & Désilets, A., Transfer Optimization in a Transit Network, *Transportation Science*, Vol. 26, No. 2, pp. 106-118, 1992.

[3] Braker, J.G., *Algorithms and Applications in Timed Discrete Event Systems*, PhD Thesis, Faculty of Technical Mathematics and Informatics, Delft University of Technology, Delft, 1993.

[4] Braker, J.G. & Olsder, G.J., The Power Algorithm in Max Algebra, *Linear Algebra and Its Applications*, Vol. 182, pp. 67-89, 1993.

[5] Cohen, G., Dubois, D., Quadrat, J.-P. & Viot, M., A Linear-System-Theoretic View of Discrete-Event Processes and Its Use

for Performance Evaluation in Manufacturing, *IEEE Transactions on Automatic Control*, Vol. AC-30, No. 3, pp. 210-220, 1985.

[6] Domschke, W., Schedule Synchronization for Public Transit Networks, *OR Spektrum*, Vol. 11, pp. 17-24, 1989.

[7] Goverde, R.M.P., Bovy, P.H.L. & Olsder, G.J., The Max-Plus Algebra Approach to Transportation Problems, paper presented at the 8th World Conference on Tranport Research (WCTR 98), Antwerp, July, 1998.

[8] Karp, R.M., A Characterization of the Minimum Cycle Mean in a Digraph, *Discrete Mathematics*, Vol. 23, pp. 309-311, 1978.

[9] Migom, A. & Valaert, W., The Optimization of a Rhytmic Intercity Train Network, *Rail International*, Vol. 12, pp. 71-91, 1981.

[10] Odijk, M.A., A Constraint Generation Algorithm for the Construction of Periodic Railway Timetables, *Transportation Research B*, Vol. 30, No. 6, pp. 455-464, 1996.

[11] Weigand, W., The Man-Machine Dialogue and Timetable Planning, *Rail International*, Vol. 14, pp. 8-25, 1983.