THE MAX-PLUS ALGEBRA APPROACH TO TRANSPORTATION PROBLEMS

ROB M.P. GOVERDE
Delft University of Technology
Faculty of Civil Engineering and Geo Sciences
Traffic Engineering and Transportation Planning Section
Stevinweg 1, P.O. Box 5048, NL-2600 GA Delft, The Netherlands

PIET H.L. BOVY
Delft University of Technology
Faculty of Civil Engineering and Geo Sciences
Traffic Engineering and Transportation Planning Section
Stevinweg 1, P.O. Box 5048, NL-2600 GA Delft, The Netherlands

GEERT JAN OLSDER
Delft University of Technology
Faculty of Information Technology and Systems
Applied Analysis Section
Mekelweg 4, NL-2628 CD Delft, The Netherlands

Abstract

This paper shows the merits of the so-called max-plus algebra as a mathematical modelling framework for Discrete Event Dynamic Systems (DEDS) in the transportation field. An overview is given of the modelling and analysis concepts of the max-plus algebra approach in the setting of a railway system. The analysis of the dynamic behaviour of transportation systems include aspects such as periodicity, stability and robustness. Also essential dynamic characteristics can be quantified such as minimum cycle time and critical circuits. As an example of the potential of this approach an analysis is given of the dynamic properties of the Netherlands intercity railway system.

INTRODUCTION

Many transportation phenomena can be described as so-called discrete event dynamic systems (DEDS). A DEDS is a dynamic asynchronous system where the state transitions are initiated by events that occur at discrete instants of time. An event corresponds to the start or the end of an activity. Examples are scheduled transportation services on networks (Braker, 1993), or the timing of a system of coordinated traffic signals along an urban arterial or in an urban road network (Egmond et al., 1999). A common property of such examples is that the start of an activity (train departure, start of green time phase at the entry of a crossing) depends on the termination of several other activities (train arrivals, platoon arrivals at a crossing). Such systems cannot conveniently be described by differential or difference equations, and naturally exhibit a periodic behaviour.

Typical questions to be answered in DEDS are whether or not the system is stable, the choice of optimal periods (scheduling), the determination of minimum (buffer) times between events, the determination of critical links or critical circuits in a network, and the quantification of sensitivity to perturbations.

Many frameworks exist to study DEDS. Examples are queueing theory, Petri nets, computer simulation, and max-plus algebra. Up to now the most widely used technique to analyse DEDS is computer simulation. An important drawback of computer simulation is that it often does not give a real understanding of the impacts of parameter changes on important system properties such as robustness, stability and optimality of system performance. Analytical techniques can provide a much better insight in such properties. Therefore, mathematical models are to be preferred as tools for modelling, analysis and control of DEDS. Cassandras et al. (1995) give an introduction to DEDS including the max-plus algebra approach.

The max-plus algebra is comparable to linear algebra. In max-plus algebra the addition (+) and multiplication ($\times$) operators from linear algebra are replaced by the maximization (max) and addition (+) operators respectively. Using these operators a linear description (in the max-plus algebra) of a nonlinear system (in the conventional algebra) is achieved and consequently a lot of concepts and properties of the classical linear system theory (eigenvectors, eigenvalues, stability, etc.) can fruitfully be exploited to study particular phenomena of DEDS. Baccelli et al. (1992) is the standard reference to the max-plus algebra theory.

This paper shows the merits of max-plus algebra as a mathematical modelling framework for DEDS in the transportation field. An overview is given of the modelling and analysis concepts of the max-plus algebra approach in the setting of a railway system, see also Goverde (1996). It is shown how essential dynamic characteristics of transportation systems can be found such as minimum cycle times, critical circuits, and stability (these concepts are defined later). As an example of the potential of this approach an analysis is given of the dynamic properties of the Netherlands intercity railway system due to Braker (1993). It is shown that the system is stable and the degree of stability is quantified. Moreover, critical lines are indicated explicitly which are crucial in the propagation of delays and the design of faster timetables.

MODELLING OF A RAILWAY SYSTEM

Example Railway System without Timetable

Consider a metropolitan area with two railway stations, S_1 and S_2, which are interconnected by a railway system as indicated in Figure 1. This railway system consists of inner circle lines and two outer circle lines. The trains on the outer circle lines deliver and pick up passengers in the suburbs. The stations in the suburbs have not been drawn since they do not play any role in the model to be formulated.

Assume there are four trains (two at each station) which leave the stations at time 0, one along each track. They reach the other (or the same) station after a certain time, the running time, which is indicated in Figure 1. A holding policy is applied, i.e., the arriving trains at a station have to wait for each other to allow passengers to change trains. The transfer time has been incorporated in the running time. After the transfer the trains leave immediately, i.e., there is no timetable. If this process of departing and arriving trains is continued, the departure time $x_i(k+1)$ for the k+1-st departure at station S_i satisfies

$$\begin{aligned} x_1(k+1) &= \max(x_1(k)+2, x_2(k)+5), \\ x_2(k+1) &= \max(x_1(k)+3, x_2(k)+3), \end{aligned} \tag{1}$$

for $k = 0,1,\ldots$. The counter k is the period indicator, i.e., from all stations the k-th train of any line leaves in period k.

Assume that $x_1(0) = 0$, $x_2(0) = 0$ are the initial conditions (initial departure times). Then the subsequent departure times from the two stations turn out to be

$$x(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \to x(1) = \begin{pmatrix} 5 \\ 3 \end{pmatrix} \to x(2) = \begin{pmatrix} 8 \\ 8 \end{pmatrix} \to x(3) = \begin{pmatrix} 13 \\ 11 \end{pmatrix} \to x(4) = \begin{pmatrix} 16 \\ 16 \end{pmatrix} \to \cdots$$

Note that the departure times follow the same pattern after each 2 periods, that is, the interdeparture times from the stations are alternately 3 and 5 time units. The average interdeparture time is 4 time units. From a timetable point of view (timetables must be as regular as possible), more convenient initial departures are $x_1(0) = 1$, $x_2(0) = 0$, since then the evolution becomes

$$x(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \to x(1) = \begin{pmatrix} 5 \\ 4 \end{pmatrix} \to x(2) = \begin{pmatrix} 9 \\ 8 \end{pmatrix} \to x(3) = \begin{pmatrix} 13 \\ 12 \end{pmatrix} \to x(4) = \begin{pmatrix} 17 \\ 16 \end{pmatrix} \to \cdots$$

where the interdeparture time is now exactly 4 at each station and the departure times are thus very

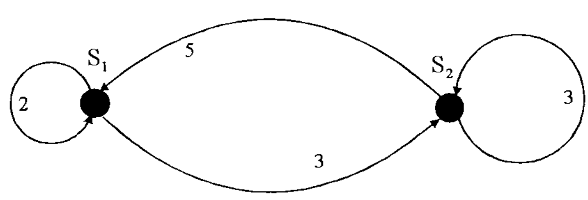

Figure 1 – Example railway network

regular (they follow the same pattern after each period). By trial and error it turns out that, whatever the initial condition, after possibly a short transient period of time, a periodic behaviour of either 1 or 2 periods is obtained with (average) interdeparture times 4. A solution with an (average) departure time smaller than 4 is not possible, since the train circulation time of the inner circle is 3+5=8 time units. There are two trains on the inner circle and therefore the (average) interdeparture time is limited from below by 8/2=4 time units.

An alternative way of writing (1) is

$$x(k+1) = A \otimes x(k), \tag{2}$$

where the state vector x consists of the two components x_1 and x_2, and $x(k)$ thus consists of the k-th departure times from all stations i, and where the running times are assembled in a matrix A given as

$$A = \begin{pmatrix} 2 & 5 \\ 3 & 3 \end{pmatrix}. \tag{3}$$

The purpose of this way of writing is that now (1) looks like a linear difference equation. The only difference of (1) with a conventional linear difference equation is that the conventional addition has been replaced by maximization and the conventional multiplication has been replaced by addition.

The vector $x(0) = (1\ 0)'$ can be interpreted as an eigenvector of A, since $x_i(1) = x_i(0) + 4$, (i=1,2), which in the '$\otimes$' notation would read $x(1) = \lambda \otimes x(0) = A \otimes x(0)$, where λ can be interpreted as an eigenvalue, with value 4,

$$x(1) = \begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = A \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = A \otimes x(0).$$

This is explained in more detail in the analysis section.

Formalization

In order to facilitate the quantitative analysis of arbitrary transport service networks this section generalizes and formalizes the above example. The basic form of the systems under consideration is

$$\begin{aligned} x_i(k+1) &= \max(a_{i1} + x_1(k), a_{i2} + x_2(k), \ldots, a_{in} + x_n(k)) \\ &= \max_{j=1,\ldots,n} (a_{ij} + x_j(k)) \qquad i = 1, \ldots, n, \end{aligned} \tag{4}$$

where a_{ij} corresponds to the running time from node j to i. Addition + will be written as $\otimes$ and max will be written as $\oplus$. This change of notation visualizes the resemblance with conventional linear difference systems:

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

which in vector notation will be written as (2), where now $x \in \mathrm{IR}^n$ and A is an $n \times n$ matrix of running times between transfer stations. One speaks of a linear (difference) equation in the max-plus algebra, this in clear analogy with linear difference equations in the conventional, 'plus-times' algebra. If it is clear from the context that the underlying algebra is the max-plus algebra then (2) is even written as $x(k+1) = Ax(k)$.

If the initial condition for (2) is $x(0) = x_0$ then $x(1) = A \otimes x_0$, and

$$x(2) = A \otimes x(1) = A \otimes (A \otimes x_0) = (A \otimes A) \otimes x_0 = A^2 \otimes x_0.$$

Here $A \otimes A$ is simply written as A^2. In general

$$x(k) = \underbrace{(A \otimes A \otimes \cdots \otimes A)}_{k \text{ times}} \otimes x_0 = A^k \otimes x_0.$$

The matrices $A^2, A^3, \ldots$ can be calculated directly. As an example, consider the A-matrix of (3). Then

$$A^2 = \begin{pmatrix} \max(2+2,5+3) & \max(2+5,5+3) \\ \max(3+2,3+3) & \max(3+5,3+3) \end{pmatrix} = \begin{pmatrix} 8 & 8 \\ 6 & 8 \end{pmatrix}.$$

In general

$$(A^2)_{ij} = \bigoplus_{l=1}^{n} a_{il} \otimes a_{lj} = \max_{l=1,\ldots,n} (a_{il} + a_{lj}). \tag{5}$$

In terms of the railway example, the quantity $(A^2)_{ij}$ can be interpreted as the maximum of all running times from station S_j via one station S_l to station S_i. One speaks of paths of length two between the stations S_j and S_i. In graph-theory terminology, the stations are called nodes and the tracks between stations are called arcs. More generally, $(A^k)_{ij}$ denotes the maximum running time of all paths of length k, starting at node j and ending at node i.

In many networks such as a railway net there will not be an arc from each node to each other node. If there is no arc from node S_j to node S_i then the departure times from node S_i are not directly influenced by those from node S_j. In such a situation it is useful to consider the element a_{ij} to be equal to $-\infty$. In the maximization of (4) a term $-\infty + x_j(k)$ does not influence $x_i(k+1)$ as long as $x_j(k)$ is finite. Minus infinity will occur frequently in the sequel and it will be indicated as $\varepsilon := -\infty$.

In practice the situation may occur that the departure time of a train from a certain station depends on a train from another station of more than one period before. Therefore, a generalization of (2) seems

$$x(k+1) = A_1 x(k) \oplus \cdots \oplus A_{l+1} x(k-l), \tag{6}$$

which is a so-called l+1-order system (note that the original system (2) equals (6) with l=0, the corresponding system is a 1st order system). However, (6) can be rewritten as a set of first order difference equations by augmenting the state space. This is a standard trick in system theory.

Railway System with Timetable

In practice a railway system operates on the basis of a timetable. A timetable contains all scheduled train departure times from all stations. This implies that a train may not leave a station before its scheduled departure time, even if all feeder trains have arrived. A generalization of the dynamic equation (2) is easily obtained to incorporate a timetable.

Let $d(k)$ be the vector of the scheduled k-th departure times of all n train services. Then (4) is extended to

$$x_i(k+1) = \max\left(a_{i1} + x_1(k), a_{i2} + x_2(k), \ldots, a_{in} + x_n(k), d_i(k+1)\right) \qquad i = 1, \ldots, n. \tag{7}$$

The k+1-th departure time of train i is thus the scheduled departure time unless a feeder train is delayed, in which case the train waits for the delayed train. In vector notation (7) is written as

$$x(k+1) = Ax(k) \oplus d(k+1), \tag{8}$$

which is a generalization of (2).

In the case of a cyclic timetable with cycle time T, i.e., all trains are scheduled modulo T, the scheduled departure times are given by

$$d(k) = d(0) \otimes T^k,$$

where T^k corresponds to kT in the conventional algebra. Of course, the timetable has to be realistic, i.e., if there are no delays then the trains should be able to operate according this timetable. A timetable is called realistic if for all $k \geq 0$

$$A \otimes d(k) \leq d(k+1).$$

The corresponding system is called a realistic system.

GRAPH THEORY

This section recapitulates the required graph theory concepts that facilitate (insight into) the algebraic definitions and theorems of the next section.

It is here assumed that A is a square matrix of which the entries may have the 'value' ε.

Definition 1 [Precedence graph] *The precedence graph $G(A)$ of an $n \times n$ matrix A is a weighted digraph with n nodes and an arc (j,i) if and only if $a_{ij} \neq \varepsilon$. The weight corresponding to an arc (j,i) is a_{ij}.*

As an example, Figure 1 is the associated precedence graph of the A matrix of (3) corresponding to the (first order) max-plus algebra model (2).

Clearly, any weighted digraph $G = (V,E)$, where V is the set of nodes and E the set of arcs, is the precedence graph of an appropriately defined square matrix. The weight a_{ij} of the arc from node j to node i is defined as the ij-th entry of a matrix A. If an arc does not exist then the corresponding entry of A becomes ε . The matrix A thus defined has G as its precedence graph.

The fact that a_{ij} corresponds to an arc (j,i) (instead to an arc (i,j)) is motivated by the interpretation of the matrix entries as running times. If $x(k)$ denotes the vector of departure times then the next departure times are given by $x(k+1)$. From (4) it follows that the departure time $x_i(k+1)$ from node i depends on the departure time $x_j(k)$ from any node j and the running time a_{ij} from node j to node i. Viewed this way, both the interpretation of the system evolution and the direction of the arcs are natural.

Let $G = (V,E)$ be a weighted digraph with n nodes. The weights form the corresponding $n \times n$ matrix A. As discussed before, the element (i,j) of $A^k = A \otimes \cdots \otimes A$, considered in the max-plus algebra, denotes the maximum weight with respect to all paths of length k from node j to node i. If no such path exists then $(A^k)_{ij} = \varepsilon$. The weight $|\rho|_{\mathrm{w}}$ of a path ρ is the (conventional) addition of the weights of all arcs constituting the path, and the length $|\rho|_{\mathrm{l}}$ of a path is the number of its arcs.

Definition 2 *The* mean weight of a path *is defined as the sum of the weights of the individual arcs of this path, divided by the length of this path. If the path is denoted by* ρ *then the mean weight equals* $|\rho|_{\mathrm{w}}/|\rho|_{\mathrm{l}}$. *If the path is a circuit then this ratio is called the mean weight of the circuit, or simply the* cycle mean.

As an example, assume that a circuit ζ is given as the sequence of nodes $i_1 \to i_2 \to \cdots \to i_{l-1} \to i_l = i_1$ then its cycle mean is

$$\frac{|\zeta|_{\mathrm{w}}}{|\zeta|_{\mathrm{l}}} = \frac{a_{i_1 i_2} + \cdots + a_{i_{l-1} i_l}}{l-1}.$$

Of interest is the maximum of these cycle means, where the maximum is taken over all circuits in the graph. This number will be called the *maximum cycle mean*. If the cycle mean of a circuit equals the maximum cycle mean then the circuit is called critical. The graph consisting of the critical circuit(s) is called the *critical graph* and denoted by G^{c}. The following definition uses the notion *strongly connected* (di-) graph. A graph is called strongly connected if there exists a path from any node to any other node. The matrix corresponding to a strongly connected graph is called *irreducible*.

Definition 3 [Cyclicity of a graph] *The cyclicity of a strongly connected graph equals the greatest common divisor of the lengths of all its circuits. The cyclicity of an arbitrary graph equals the least common multiple of the cyclicities of all its maximal strongly connected subgraphs.*

The next section considers the correspondence of the cyclicity of a graph with the concept of cyclicity of a matrix.

ANALYSIS

This section considers the powerful analysis concepts which are the main advantage of the max-plus algebra approach. Important concepts are the periodicity or cyclicity of a scheduled transport service system, the minimum cycle time, the critical circuit, stability, and sensitivity to perturbations (delays). To improve the insight in the algebraic properties, the correspondence to their graph-theoretic counterparts is considered where possible.

Periodicity

A transportation system naturally has a periodic behaviour in the sense that a pattern of departure times of all train services repeats itself after a particular number of periods, the cyclicity. This periodicity or cyclicity is the basis of a cyclic timetable.

Definition 4 [Cyclicity of a matrix] *A matrix A is said to be cyclic if there exist scalars M, λ and d such that $\forall m \geq M$, $A^{m+d} = \lambda^d A^m$. The least such d is called the cyclicity of A. The quantity λ equals the maximum cycle mean of A.*

The expression $A^{m+d} = \lambda^d A^m$ in the definition above must be interpreted in the max-plus algebra sense. Hence λ^d in the max-plus algebra means $d\lambda$ in the conventional algebra and $\lambda^d A^m$ refers to the addition of λ^d to each element of A^m.

In terms of the railway example a cyclicity d implies that irrespective of the initial departure times x_0 a periodic behaviour of (at least) d periods is obtained with average interdeparture time λ (possibly after a short transient period of time). If a matrix has cyclicity d then A is also called order-d periodic.

The following theorem gives a method to find the cyclicity of a matrix by using the graph-theoretic counterpart (Baccelli et al., 1992).

Theorem 1 *Any irreducible matrix is cyclic. The cyclicity of the irreducible matrix A equals the cyclicity of $G^c(A)$, the critical graph corresponding to matrix A.*

Example Consider matrix A of (3). The corresponding precedence graph (see Figure 1) has three circuits, viz. from node 1 to node 1 with cycle mean 2/1=2; from node 1 via 2 to 1 with cycle mean (5+3)/2=8/2=4; from node 2 to 2 with cycle mean 3/1=3. (It is tacitly assumed here that node i corresponds to x_i.) The maximum cycle mean hence equals 4, and the corresponding (unique) critical circuit is the inner circle. The critical graph consists of the unique critical circuit only. The cyclicity of the critical graph equals 2, since the critical circuit consists of 2 arcs. The quantities of Definition 4 are $M = 2$, $\lambda = 4$ and $d = 2$,

$$A^4 = \begin{pmatrix} 16 & 16 \\ 14 & 16 \end{pmatrix} = 4^2 \otimes A^2 = 8 \otimes \begin{pmatrix} 8 & 8 \\ 6 & 8 \end{pmatrix}.$$

Note that a scalar multiplication of a matrix in the max-plus algebra equals the addition of the scalar to each matrix entry in the conventional algebra. It was already noted that for the initial departure times $x_i(0)=0$, (i=1,2), an order-2 periodic pattern resulted with an average interdeparture time of 4 (the maximum cycle mean).

Minimum cycle time

Let A be a matrix in the max-plus algebra. Consider the problem of finding an eigenvalue λ and eigenvector $v \neq \varepsilon$ such that

$$Av = \lambda v. \tag{9}$$

This equation has to be interpreted in the max-plus algebra sense: the expression λv means that one adds λ to each component of v. If the eigenvector v corresponds to an initial state then this results in a solution with an order-1 periodic pattern and λ is the interdeparture time.

The following theorem gives the correspondence between the eigenvalue of a (square) matrix and the maximum cycle mean of its precedence graph (Baccelli et al., 1992).

Theorem 2 *Let A be a square matrix. If $G(A)$ is strongly connected then there exists a unique eigenvalue and at least one eigenvector. The eigenvalue is equal to the maximum cycle mean of the graph,*

$$\lambda = \max_{\zeta} \frac{|\zeta|_w}{|\zeta|_1},$$

where ζ ranges over the set of circuits of $G(A)$.

The eigenvalue, or equivalently the maximum cycle mean, is the minimum cycle time for which a realistic timetable exists. From a graph-theoretic view the eigenvalue is the average interdeparture time of a train on the critical circuit. Viewed this way, the critical circuit is the 'slowest' circuit in the network. The mean cycle time of this circuit determines the minimum cycle time for the entire system.

An eigenvector v determined by (9) consists of the departure times of all trains such that the departure times in the next period, determined by Av, are as early as possible (irrespective of a timetable). The corresponding minimum interdeparture time is given by the eigenvalue, as follows from the right-hand side of (9). If the initial departure times are determined by the eigenvector, $x(0)=v$, then all subsequent departure times (in the absence of delays) are given as

$$x(k) = \lambda^k v,$$

which is the most regular possible behaviour for the system. In practice however, it is not advisable to operate according a timetable with minimum cycle time. A timetable should be able to compensate for slight irregularities (delays) in running times. Therefore, margins and buffer times have to be incorporated in the timetable, which increases the timetable cycle time (Goverde, 1998a). This is the subject of the stability section.

There are efficient algorithms to compute the eigenvalue, eigenvector and the critical circuit. Braker and Olsder give a power algorithm (Braker and Olsder, 1993; Braker, 1993) for computing the eigenvalue and eigenvector. The computational complexity depends on the transient behaviour of the system. Karp gives a polynomial algorithm to compute the eigenvalue and critical circuit (Karp, 1978). Howard's algorithm is a fast algorithm for computing the eigenvalue and eigenvector that has an average linear computation time (Cochet-Terrasson et al., 1998).

Stability

An important issue in transportation systems is the sensitivity to delays that occur during operation. The transportation system should be able to cope with a certain amount of irregularity (delays) without severe consequences. This leads to the subject of stability and robustness.

A scheduled transport service system is stable if any initial delay disappears after a finite number of periods (subsequent departures). The following theorem gives an elegant method to test stability (Braker, 1993).

Theorem 3 [Stability] *The system with timetable (8) is stable if and only if*

$$\lambda < T,$$

where T is the (timetable) cycle time and λ is the eigenvalue (or maximum cycle mean).

From Theorem 3 it follows that the cycle time T of the timetable must be larger than the eigenvalue to obtain a stable system, which corresponds to the interpretation of the eigenvalue as minimum cycle time.

Stability of a scheduled transport service system only implies that any delay will eventually disappear. However, it is also worthwhile to know how long this might take. In practice, scheduled transfer times between two connected trains at a transfer station include buffer times, i.e., a transfer time consists of a minimum transfer time in which transferring passengers are able to get to the connecting train (with high probability) and a transfer buffer time. The transfer buffer time is due to network constraints in the timetable design process, and/or deliberately incorporated to compensate for arrival delays and thereby increasing the transfer reliability.

Using the max-plus algebra model, the propagation of any initial delays is easily computed. Define the initial state as the vector of which each component is the sum of the scheduled departure time and the initial delay. The evolution of the system can then be computed by the dynamic equation (8). As soon as all computed departure times equal the scheduled departure times the delays have disappeared. The corresponding period is called the settling time corresponding to the particular initial delay. Note that the effect of the delay to subsequent departure times of all trains is computed explicitly. It is here tacitly assumed that all trains wait for delayed feeder trains. An extension is that planned transfers may be cancelled when feeder trains are behind schedule. This control of connections can also be incorporated in the max-plus algebra model (Braker, 1993). This delay propagation model efficiently computes secondary waiting times in a network which can be used in the derivation of optimal train waiting times at transfer stations to secure connections (Goverde, 1998b).

The following definition gives an analytical tool to test the sensitivity (or robustness) of the system with respect to the transfer times.

Definition 5 [Stability margin] *The* stability margin Δ *of a stable system* (8) *is the minimum amount of additional transfer time for which it is impossible to find a timetable, with the same cycle time T, under which the system is stable.*

A lower bound for Δ is given by $T-\lambda$. The fact that $\lambda < T$ thus shows that the system is stable, but the value $T-\lambda$ gives information on how stable. Clearly, the gap between the minimum cycle time λ and the timetable cycle time T is a measure for the sensitivity of the system. Increasing the transfer times at all nodes in the critical circuit by $T - \lambda$ results in a circuit with maximum cycle mean that equals or is lower than T. The timetable is therefore still stable.

However, the stability margin might be larger than this lower bound. If the 1st order system (8) has been obtained from the original description (6) by augmenting the state space, i.e., by introducing auxiliary nodes, a different situation occurs. The auxiliary variables can be viewed as dummy nodes which naturally have no transfer times. If the critical circuit contains dummy nodes then more transfer time can be added to the 'real' nodes before the cycle mean of this circuit equals T. Moreover, another circuit may become critical by increasing the transfer times. This may occur when this other circuit has fewer dummy nodes.

An upper bound for the stability margin is given by

$$\Delta \leq (T - \lambda)\frac{m}{m - \widetilde{m}}, \tag{10}$$

where m is the number of nodes in the critical circuit and $\widetilde{m}$ is the number of auxiliary nodes in the critical circuit (Braker, 1993). Braker (1993) gives an algorithm to compute the exact value of the stability margin.

THE NETHERLANDS INTERCITY RAILWAY EXAMPLE

This section applies the introduced max-plus algebra notions to the analysis of the intercity train service network in the Netherlands. It also considers questions related to the timetable design corresponding to the underlying network. The intercity network is assumed to be a closed railway network. The results are due to Braker (1993) and correspond to the railway timetable of 1989/1990.

The Netherlands intercity network consists of 11 lines (routes along which trains run in both directions), see Figure 2. As an example, line 10 starts (respectively terminates) at Amsterdam and terminates (respectively starts) at Vlissingen. Trains arriving at both end points turn around and continue in opposite direction. Each line is thus a circuit. A fixed number of trains runs on each line. The train running times between stations are given and assumed to be deterministic. International connections are not considered: the model assumes that trains arriving at the border turn around and continue in opposite direction (which actually happens with some trains at some border stations). Hence the name 'closed' network. The timetable has a cycle time of 30 min.

Assume there is no timetable. In addition to the line structures and running times between stations also realistic connection constraints are taken into account to allow passengers to transfer. These connection constraints give for each station the set of trains that have to wait for each other to guarantee a transfer. In total, there are 53 trains that have scheduled transfers where different lines meet. As an example, the train of line 10 that has a scheduled departure time at Leiden at time instant 00 in the direction Den Haag has two connection constraints. It has to wait on the train from Amsterdam of line 10 with scheduled departure time 26 and running time 34 min, and on the train from Haarlem of line 41 with scheduled departure time 10 and running time 18 min. The transfer time is here incorporated in the running times. Note that the train at Leiden is in fact physically the same train as the one arriving from Amsterdam. The connection is here a stop connection. Finally, also setup times at end points are taken into account.

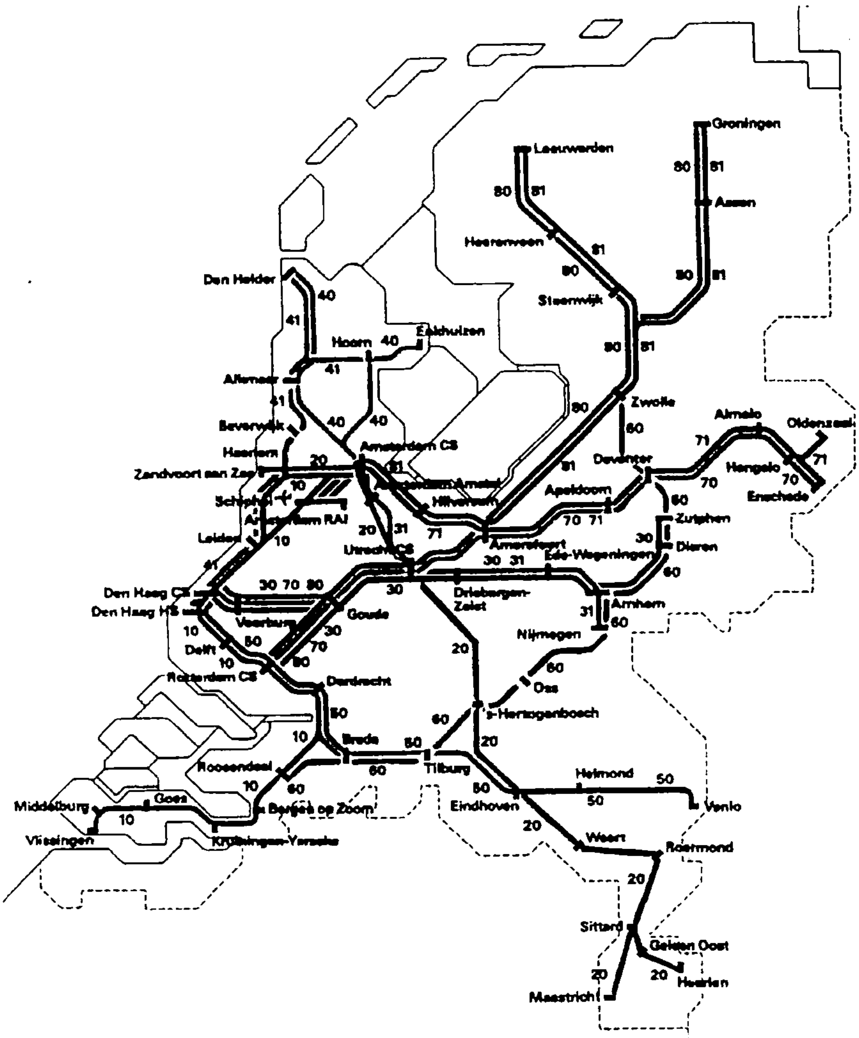

Figure 2 - Dutch intercity network

The resulting max-plus model has the form (6), where x is a 53-dimensional vector, corresponding to the 53 trains with connection constraints. After rewriting (6) as a 1st-order model (2), a model with state vector of dimension 79 results. The running time matrix A is not given here: it is made up of the various train running times as given in the published timetable, and standard transfer times as defined by the railway company. The model can now be used to analyse the propagation of delays on the intercity network and to compute the corresponding settling time. Moreover, the max-plus model offers interesting analysis tools such as the eigenvalue, critical circuit and stability margin which is the major advantage of the max-plus modelling.

The calculated critical circuit (which is not to be confused with a line of the network) turns out to go from Venlo via Eindhoven, Utrecht and Amsterdam to Zandvoort and back. The circuit consists of 12 nodes of which 6 are auxiliary nodes, and the circulation time of the circuit is 325 minutes.

This circuit is the bottleneck for a faster timetable. Hence, if a faster or more robust timetable is desired trains on this circuit have to 'speed up'. This can be accomplished by, for instance, adding extra trains to a line (or lines) which forms part of this critical circuit (i.e., the lines numbered 20 and 50), by infrastructural investments to shorten relevant track section running times, or by redesigning the line structure.

The weight of the critical circuit is 325 minutes, and the length is 12. The maximum cycle mean corresponding to the critical circuit, or equivalently the eigenvalue, is thus $\lambda = 325/12 = 27\frac{1}{12}$. As the actual timetable cycle time is T = 30 minutes, by Theorem 3 follows that the railway system is stable. The difference between the actual 30 min and the theoretical λ is the flexibility in the system which causes propagation of possible delays to disappear in finite time.

The lower bound for the stability margin Δ is given by $T - \lambda = 30 - 27\frac{1}{12} = 2\frac{11}{12}$. The upper bound can be calculated by (10). Here, the number of nodes in the critical circuit is m = 12 and the number of auxiliary nodes is $\tilde{m} = 6$. It follows that

$$\Delta \leq (30 - 27\tfrac{1}{12})\frac{12}{12-6} = 5\tfrac{5}{6}$$

The actual stability margin turns out to be $\Delta = 5\frac{2}{3}$. If this number is added to all transfer times, to be incorporated into the model by adjusting the A matrix, then the eigenvalue is $\lambda = 30$ minutes. The corresponding system will then have maximum transfer times based on a timetable with a cycle time T = 30 min. The critical circuit, with cycle mean $\lambda = 30$, is now the circuit from Venlo via Eindhoven, Utrecht, Amsterdam, Haarlem, Den Haag HS, Breda and Eindhoven back to Venlo.

For larger delays a strategy might be not to wait for delayed feeder trains. A train thus only stops so that passengers are able to alight and board, and if the train is already behind schedule it no longer waits for feeder trains but leaves immediately. For this controlled model a faster timetable exists with $\lambda = 26\frac{2}{11}$. The critical circuit is now the circuit from Venlo via Eindhoven and Breda to Den Haag CS and back.

CONCLUSIONS

Transportation systems can be modelled rather naturally as a DEDS using the max-plus algebra. The nonlinear behaviour of the synchronization processes can be written as a linear system (in the max-plus algebra). A straight correspondence with graph-theoretic concepts with algebraic ones results in a clear visualization of the max-plus algebra model and insight into the model properties. Moreover, the parallels between max-plus algebra system theory and conventional linear system theory offer elegant analysis concepts, such as the eigenvalue and the eigenvector.

A case study of the analysis of the Netherlands intercity railway system has illustrated the potential of the max-plus algebra approach to relevant and large-scale problems. Subiono (1997) has extended the results to the entire Dutch railway network including intercity (IC) trains, interregional (IR) trains and agglo-regional (AR) trains.

The max-plus algebra model as considered in this paper does not take into account stochastic train running times and infrastructure limitations (minimum headways). Incorporating these concepts in the model to obtain a more realistic model is the subject of current research. However, a model is always a simplification of reality, and although a more realistic model might be necessary for, for instance, a reliable simulation model, the considered model is sufficient for detecting bottlenecks and for stability analysis with respect to (intercity) train connections.

This paper showed the applicability of the max-plus algebra modelling to railway service networks. Egmond et al. (1999) applies the max-plus algebra approach to the synchronization of traffic signals in an urban (road) network. Other potential application fields include air traffic planning, freight transport planning, and planning of unimodal and intermodal connections in multimodal systems.

ACKNOWLEDGEMENTS

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

REFERENCES

Baccelli, F.L., Cohen, G., Olsder, G.J., Quadrat, J.-P. (1992) **Synchronization and Linearity: an Algebra for Discrete Event Systems**. Wiley, Chichester.

Braker, J.G. (1993) **Algorithms and Applications in Timed Discrete Event Systems**. PhD Thesis, Faculty of Information Technology and Systems, Delft University of Technology, Delft.

Braker, J.G. and Olsder, G.J. (1993) The Power Algorithm in Max Algebra. **Linear Algebra and Its Applications 182**, 67-89.

Cassandras, C.G., Lafortune, S., Olsder, G.J. (1995) Introduction to the Modelling, Control and Optimization of Discrete Event Systems. In: Isidori, A. (ed.), **Trends in Control: a European Perspective**, 217-291. Springer Verlag, Berlin.

Cochet-Terrasson, J., Cohen, G., Gaubert, S., Gettrick, M. Mc., Quadrat, J.-P. (1998) Numerical Computation of Spectral Elements in Max-Plus Algebra. **Paper of the IFAC Conference on System Structure and Control**, Nantes, France, 8-10 July, 1998.

Egmond, R.-J. van, Olsder, G.J., Zuylen, H.J. van (1999) A New Way to Optimise Network Traffic Control Using Maxplus Algebra. **Proceedings of the TRB 78th Annual Meeting**, Washington, 10-14 January, 1999.

Goverde R.M.P. (1996) **Max-Plus Algebra voor de spoorwegen**. Report, TRAIL Research School, Delft (in Dutch).

Goverde, R.M.P. (1998a) The Max-Plus Algebra Approach to Railway Timetable Design. In: Mellitt, B., Hill, R.J., Allan, J., Sciutto, G., Brebbia, C.A. (eds.), **Computers in Railways VI**, 339-350. Computational Mechanics Publications / WIT Press, Southampton.

Goverde, R.M.P. (1998b) Synchronization Control of Scheduled Train Services to Minimize Passenger Waiting Times. **Proceedings of the 4th TRAIL Year Congress 1998**, Part 2, TRAIL Research School, Delft.

Karp, R.M. (1978) A Characterization of the Minimum Cycle Mean in a Digraph. **Discrete Mathematics 23**, 309-311.

Subiono (1997), On Large Scale Max-Plus Algebra Models in Railway Systems. **Proceedings of the 3rd TRAIL Year Congress 1997**, TRAIL Research School, Delft.