Safeness-Enforcing Supervisory Control for Railway Networks

Fabrizio Diana, Alessandro Giua, Carla Seatzu

Department of Electrical and Electronic Engineering, University of Cagliari
Piazza d’Armi — 09123 Cagliari, Italy
{giua,seatzu}@diee.unica.it
http://www.diee.unica.it/~giua/info.html

Abstract — In this paper we deal with the problem of modeling railway networks with Petri nets so as to apply the theory of supervisory control for discrete event systems to automatically design the system controller. We provide a modular representation of railway networks in terms of stations and tracks including sensors and semaphores. We ensure safeness and local liveness imposing both Generalized Mutual Exclusion Constraints and constraints also involving the firing vector.

I. INTRODUCTION

The specification, analysis and implementation of railway control logic has ever been an important activity since trains and railways were invented centuries ago, and failure of control logic can lead to railway accidents and loss of human life. At present time, this activity is even more important because railway networks are often large, the speed of trains and traffic density is increasing, and activities within networks are taking place concurrently and at geographically different locations. As a result, the overall complexity of railway systems increases, and hence greater demands are placed on the control logic of these systems [10].

The control of a railway network can be divided into two distinct phases. The first one, at a lower level, imposes the satisfaction of a series of safeness constraints (collision avoidance) and liveness constraints (deadlock freeness). The second one, at a higher level, is concerned with the problem of scheduling both the departures and the stops, so as to optimize the efficiency of the net. In this paper the attention is uniquely devoted to the first phase.

We focus our attention on the modeling and control of railway networks with Petri nets [14], that provide a powerful framework for the analysis and control of distributed and concurrent systems. Some of the advantages of Petri nets as models for discrete event control include [8]: graphical representation, solid foundations based in mathematics, the existence of simulation and formal analysis techniques, and the existence of computer tool support for simulation, analysis and control. The literature on modelling and analyzing railway systems using Petri nets is not extensive and a good survey is given by Janczura in [10]. The idea of applying Petri nets theory goes back to Genrich [5], then it was revisited in [1], [11] and in [9] where coloured Petri nets have been used. Significant contributions in this setting are also due to Deckmatel and Schnieder [2] and Di Febbraro et al. [4] who used hybrid Petri nets to model transportation systems.

We provide a modular representation of railway networks in terms of stations and tracks including sensors and semaphores. The overall model will be a place/transition (P/T) net whose transitions may be (un)controllable and/or (un)observable, following the paradigm of supervisory control [15]. There exist several techniques for automatically designing controllers for P/T nets with uncontrollable and/or unobservable transitions [8]. In particular, we show how collision avoidance constraints can be expressed as Generalized Mutual Exclusion Constraints (GMECs) [6] and the corresponding controller takes the form of a set of monitor places that can be automatically computed — taking into account uncontrollable and unobservable transitions — using Moody’s parametrization [13].

It is well known that a P/T net monitor based solution may not be maximally permissive when there exist uncontrollable or unobservable transitions [6]: the price one has to pay to keep the control structure simple is the fact that the controller may unnecessarily disable some transitions. In the case at hand, this leads to a local deadlock, i.e. the automatically designed monitor controller leads to a block when two trains coming from opposite directions cross each other at a station. To solve this problem we modify the controller, writing explicitly a new set of rules that define the admission policy into the stations; we show that the corresponding control structure is still very simple and takes the form of a “monitor with self-loops”.

A nice feature of this approach is that the whole control problem can be divided into a certain number of subproblems, thus making the proposed control procedure suitable even for large-scale networks.

Let us finally observe that in [7] we have also addressed the problem of global deadlock avoidance. In fact, when all the modules introduced here are put together and the number of trains in the network increases, it may well be the case that the net enters a blocking state. In [7] a solution to this problem has been provided applying siphon analysis to a simplified net and adding new monitors that, controlling the net siphons to prevent them from becoming empty, ensure global liveness for the system.
II. BACKGROUND

A. Generalities on Petri nets

In this subsection we recall the formalism used in the paper. For more details on Petri nets we address to [14].

A Place/Transition net \( P/T \) net is a structure \( N = (P, T, Pre, Post) \), where \( P \) is a set of \( m \) places; \( T \) is a set of \( n \) transitions; \( Pre: P \times T \to N \) and \( Post: P \times T \to N \) are the \( Pre- \) and \( Post- \) incidence functions that specify the arcs; \( C = Post - Pre \) is the incidence matrix.

A marking is a vector \( m: P \to N \) that assigns to each place of a \( P/T \) net a non-negative integer number of tokens, represented by black dots. In the following we denote as \( m_i \) the marking of place \( p_i \).

A \( P/T \) system or net system \( (N, m_0) \) is a net \( N \) with an initial marking \( m_0 \).

A transition \( t \) is enabled at \( m \) if \( m \geq Pre(\cdot, t) \) and may fire yielding the marking \( m' = m + C(\cdot, t) \). The notation \( m(t)m' \) means that an enabled transition \( t \) may fire at \( m \) yielding \( m' \).

A firing sequence from \( m_0 \) is a (possibly empty) sequence of transitions \( \sigma = t_1 \ldots t_k \) such that \( m_0(t_1m(t_1) \ldots t_km_k). \) A marking \( m \) is reachable in \( (N, m_0) \) iff there exists a firing sequence \( \sigma \) such that \( m_0(\sigma)m \). Given a net system \( (N, m_0) \) the set of reachable markings is denoted \( R(N, m_0) \). The function \( \sigma(t) \) represents the number of occurrences of \( t \) in \( \sigma \), is called firing count vector of the fireable sequence \( \sigma \). If \( m_0(\sigma)m \), then we can write in vector form \( m = m_0 + C(\cdot, t) \cdot \sigma \). This is known as the state equation of the system.

B. Generalized Mutual Exclusion Constraints

Assume we are given a set of legal markings \( L \subseteq N^n \), and consider the basic control problem of designing a supervisor that restricts the reachability set of the plant in closed loop to \( L \cap R(N, m_0) \). Of particular interest are those PN state-based control problems where the set of legal markings \( L \) is expressed by a set of \( n \) linear inequality constraints called Generalized Mutual Exclusion Constraints (GMECs).

Each GMEC is a couple \( (w, k) \) where \( w: P \to Z \) is a \( m \times 1 \) weight vector and \( k \in Z \). Given the net system \( (N, m_0) \), a GMEC defines a set of markings that will be called legal markings: \( M(w, k) = \{ m \in N^n | W^Tm \leq k \} \). The markings that are not legal are called forbidden markings. A controlling agent, called supervisor, must ensure that the forbidden markings will not be reached. So the set of legal markings under control is \( M_c(w, k) = M(w, k) \cap R(N, m_0) \).

In the presence of multiple constraints, all constraints can be grouped and written in matrix form as

\[
W^Tm \leq k
\]

where \( W \in Z^{m \times n} \) and \( k \in Z^n \). The set of legal markings is \( M(W, k) = \{ m \in N^n | W^Tm \leq k \} \).

Each constraint requires the introduction of a new place (denoted as monitor place or controller place). To each monitor place, it corresponds an additional row in the incidence matrix of the closed loop system. In particular, let \( C_c \) be the matrix that contains the arcs connecting the controller places to the transitions of the plant, and \( (m_0, m_c) \) the (initial) marking of the controller. The incidence matrix \( C \in Z^{(m+n_c) \times n} \) of the closed loop system is

\[
C = \begin{bmatrix}
C_p \\
C_c
\end{bmatrix}
\]

and the marking vector \( m \in Z^{m+n_c} \) and initial marking \( m_0 \) are

\[
m = \begin{bmatrix}
m_p \\
m_c
\end{bmatrix}, \quad m_0 = \begin{bmatrix}
m_{p0} \\
m_{c0}
\end{bmatrix},
\]

where the subscript \( p \) has been used to denote the variables of the plant.

In the case of controllable and observable transitions, Gua et al. provided the following theorem.

**Theorem 1 ([6])** If \( k - W_p^T m_0 \geq 0 \) then a Petri net controller with incidence matrix \( C_c = -W_c^T C_p \) and initial marking \( m_{c0} = k - W_p^T m_0 \) enforces constraint (1) when included in the closed loop system (2) with marking (3).

The controller so constructed is maximally permissive, i.e., it prevents only transitions firings that yield forbidden markings. The controller net has \( n_c \) control places and no transition is added.

It often occurs that certain transitions can not be disabled by any control action (uncontrollable transitions) or can not be directly detected or measured (unobservable transitions). The uncontrollability of a transition indicates that we can not add any arc from the controller places to this transition, so that the controller may never disable it. The unobservability of a transition implies that it must have the same number of input and output arcs to/from each controller place — i.e., its only admissible connection to monitor places is given by self-loops — so that its firing does not modify the controller state. In these cases the previous theorem is no more valid and an appropriate set of transformed constraints needs to be determined so as to construct a Petri net controller.

**Theorem 2 ([12])** Let a Petri net with incidence matrix \( C_p \) be given with a set of uncontrollable and/or unobservable transitions. Let \( C_{uc} \) (\( C_{uo} \)) be the incidence matrix of the uncontrollable (unobservable) portion of the Petri net. A set of linear constraints on the net marking, \( W^Tm \leq k \), are to be imposed. Assume \( R_1 \in Z^{n_{uc} \times m} \) satisfy \( R_1 m \geq 0 \) \( \forall m \in N^n \), and \( R_2 \in Z^{n_{uo} \times m} \) be a positive definite diagonal matrix, with \( R_1 + R_2 W^T \neq 0 \). Let

\[
\begin{bmatrix}
C_{uc} & C_{uo} & -C_{uo} \\
W^T C_{uc} & W^T C_{uo} & -W^T C_{uo} & W^T m_{p0} - k - 1
\end{bmatrix} \leq \begin{bmatrix}
R_1 & R_2
\end{bmatrix}
\]
we show that the whole network can be seen as the composition of a certain number of elementary modules, namely tracks and stations.

A. The track model

An example of Petri net modeling a track is shown in figure 3. It consists of two series of places (p_1, ..., p_k and p'_1, ..., p'_k) and transitions (t_1, ..., t_4 and t'_1, ..., t'_4), each one representative of the flow of trains in a certain direction. Each couple of places p_i, p'_i represents a segment of the track, i.e., the marking of either p_i or p'_i denotes the presence of a train in the segment. Note that, in the case of a double track, the two lines are independent and places p_i and p'_i correspond to parallel segments and can be marked simultaneously. On the contrary, in the case of a single track two places, p_i and p'_i, are used to represent the same segment of the track that can be crossed in both directions, but places can not be marked at the same time. Note that during simulation a release delay is associated to each transition, to represent the time a train requires to run along that segment.

Transitions may be (un) controllable and/or (un) observable. In this setting, a transition that is both controllable and observable represents a semaphore (see transitions t_2 and t'_2 in figure 3), i.e., in that point of the net the presence of a train can be detected and its transit can be forbidden. In all real situations a semaphore is placed at the exit of a track, or equivalently at the entrance of a station. A transition that is observable but not controllable (see transitions t_1, t_4, t'_1 and t'_4), represents a sensor counting the number of axles of the train, i.e., the number of cars passing through that point.

The number of places used to represent the track depends on the required precision. On one hand, we assume that the Petri net is safe (such a condition will be imposed by the addition of appropriate monitor places), thus the number of places is mainly limited by the required safeness distance, i.e., we assume that the length of each segment is such that no more than one train can be contained within it at any given time instant. On the other hand, we take into account the presence of sensors and semaphores that are modeled by appropriate transitions as discussed above. Note that, even if these elements are only associated to one direction of flow, an equal number of uncontrollable and unobservable transitions should be added in the other direction so as to keep the structure shown in figure 3. Let us finally observe that an arbitrary large number of places may be included in the model of the track, and these places should be connected through uncontrollable and unobservable transitions. Nevertheless they would only change the modeling granularity, while no variation would occur in the controller design.

B. The railway station model

In this subsection we present the Petri net model of a three-tracks railway station that is sketched in figure 4. a (ignore place p_M and all connected arcs), where double arrows have been used to denote self-loops. Note that we
Fig. 4. The Petri net model of a three-tracks railway station with the monitor place relative to the constraint $m_i + m'_i \leq 1$ assuming that all transitions are controllable and observable (a) and taking into account the uncontrollability and unobservability of transitions (b).

can easily extend this model to a railway station with an arbitrary number of tracks.

As shown in figure 4.a, the station is composed of five different stretches shown within dashed boxes: three parallel stretches in the station and input tracks on both sides. The models of the stretches are similar to those already presented in the previous subsection.

The firing of controllable and observable transitions $t_{in_1}$ and $t_{in_2}$ represent the input of a train in the station, while the firing of uncontrollable and unobservable transitions $t_{out_1}$ and $t_{out_2}$ represent the output of a train from the station. Note that, as in the case of the track model, a controllable and observable transition is used to model a semaphore, while an observable but uncontrollable transition is used to model an axles counter.

The two subnets containing places $P_{u,1}, P_{u,2}, P_{d,1}, P_{d,2}$ and $P_{u,2}, P_{d,2}, P_{i,1}$, and transitions $t_{in_1}, t_{in_2}, t_{d_1}, t_{d_2}, t_{ud_1}, t_{ud_2}$, respectively, model the points, i.e., when places $P_{u,1}$ and $P_{u,2}$ are marked, trains may be directed to the up-track or may leave the up-track; on the contrary, when places $P_{d,1}$ ($P_{i,1}$) and $P_{d,2}$ ($P_{i,2}$) are marked, trains may be directed to the down (intermediate)-track or may leave the down (intermediate)-track.

Let us finally observe that as in the case of the track model, during simulation, we associate a time delay to each transition so as to simulate time intervals required to get across a given segment of the station.

IV. THE CONTROLLER DESIGN FOR TRACKS AND STATIONS

As already discussed in the introduction, in this paper we shall deal with the problem of designing a Petri net supervisor for a railway network so as to ensure safeness and local liveness. In other words, the goal of the supervisory controller is that of guaranteeing that two trains may flow through the net in opposite directions without colliding, while prohibiting that blocking conditions may occur.

Note that as the number of trains in the system increases,
additional blocking conditions may occur. This issue has been discussed in [7].

To do this we first consider single modules and derive a controller for tracks and stations separately. In particular, we observe that GMECs may be satisfactorily applied when controlling tracks. On the contrary, this kind of constraints are too restrictive when controlling stations. We show in detail that in this latter case, safeness may be ensured by imposing appropriate logical constraints that also ensure local liveness.

GMECs have been firstly imposed so as to ensure safeness, i.e., to ensure that each couple of places corresponding to the same segment of a single-track (that may also belong to a station) are not marked simultaneously, and each place never contains more than one token at a time.

In accordance to the supervisory control theory briefly summarized in subsection II-B each constraint requires the introduction of a monitor place. Moreover, in the case of uncontrollable and/or unobservable transitions, constraints need to be appropriately transformed. As an example, a constraint of the form

$$m_i + m'_i \leq 1$$

relative to a given segment of a track within a station (see figure 4.a) ensures that places $p_i$ and $p'_i$ are not marked at the same time and each place never contains more than one token. If all transitions were controllable and observable, the monitor place $PM_i$ enforcing (5) would have been that in figure 4.a. The presence of uncontrollable and unobservable transitions requires the transformation of (5) into a more restrictive constraint

$$m_{i-3} + m_{i-2} + m_{i-1} + m_i + m'_i + m'_{i+1} + m'_{i+2} + m'_{i+3} \leq 1.$$  

(6)

The corresponding monitor place is $PM_i$ in figure 4.b. This result can be formally obtained with Moody’s procedure [13] as discussed in subsection II-B.

Constraints of this kind ensure safeness. Nevertheless, as we now show, they are too restrictive when applied to places relative to tracks within the stations, while they ensure a satisfactory behaviour of the net when imposed to places modeling the intermediate tracks.

As an example, let us consider the monitor places $PM_i$ and $PM_j$ in figure 4.b, that have been introduced so as to enforce the GMECs $m_i + m'_i \leq 1$ and $m_j + m'_j \leq 1$, respectively, and applying Moody’s transformation to make them controllable and observable. By looking at figure 4.b we can immediately observe that whenever a train is in the upper track (place $p_i$ or $p'_i$ is marked) no train can enter the station (transition $t_{m_{i-2}}$ is not enabled because place $PM_i$ is empty). Consider now the case in which a train is in the upper track going right (place $p_i$ marked) and another one is arriving from the right (place $p'_i$ marked) as shown in figure 4.b: a deadlock occurs. Such a case, as well as other analogous conditions for trains in the intermediate and lower tracks, demonstrate that GMECs do not guarantee a satisfactory behaviour of the net, resulting to be too restrictive.

A better solution to this problem consists in the introduction of a new set of constraints, some of whom may also involve the firing vector, that regulate the input of trains in the stations, and the points within them. In fact, whenever there is at least one empty stretch a reasonable...
admission policy should let the incoming train enter the station, while the points should direct the incoming train towards the empty stretch.

To do this, let us consider the simplified Petri net model of the three–tracks railway station reported in figure 5, that is obtained from the previous one by simply grouping together some places. The set of constraints that have been proved to satisfy the desired requirement of safeness and local liveness is \[3\]:

\[
\begin{align*}
  m_1 + m_2 + m_3 + m_4 + m_5 & \leq 3 \quad (a) \\
  q_{in,1} + m_2 + m_{u,1} & \leq 2 \quad (b) \\
  q_{in,2} + m_3 + m_{i,1} & \leq 2 \quad (c) \\
  q_{in,2} + m_4 + m_{d,1} & \leq 2 \quad (d) \\
  q_{u,1} + m_1 + m_2 & \leq 2 \quad (e) \\
  q_{u,1} + m_1 + m_2 & \leq 2 \quad (f) \\
  q_{i,1} + m_1 + m_2 & \leq 2 \quad (g) \\
  q_{d,1} + m_1 + m_2 & \leq 2 \quad (h) \\
  q_{d,1} + m_1 + m_2 & \leq 2 \quad (i) \\
  q_{d,2} + m_1 + m_2 & \leq 2 \quad (j) \\
  q_{in,2} + m_2 + m_{u,2} & \leq 2 \quad (k) \\
  q_{in,2} + m_2 + m_{i,2} & \leq 2 \quad (l) \\
  q_{in,2} + m_2 + m_{d,2} & \leq 2 \quad (m) \\
  q_{u,2} + m_2 + m_3 & \leq 2 \quad (n) \\
  q_{u,2} + m_2 + m_3 & \leq 2 \quad (o) \\
  q_{i,2} + m_2 + m_3 & \leq 2 \quad (p) \\
  q_{d,2} + m_2 + m_3 & \leq 2 \quad (q) \\
  q_{d,2} + m_2 + m_3 & \leq 2 \quad (r) \\
  q_{d,2} + m_2 + m_3 & \leq 2 \quad (s) \\
  q_{d,2} + m_2 + m_3 & \leq 2 \quad (t) \\
  q_{d,2} + m_2 + m_3 & \leq 2 \quad (u)
\end{align*}
\]

where constraints (a)–(d) and (m)–(o) regulate the input of trains in the station, while the others regulate the points. Note that, apart from constraint (a) that is a GMEC, all the other constraints also involve the firing vector.

The inequality (a) implies that transitions \(t_{in,1}\) and \(t_{in,2}\) may only fire if no more than two trains are already contained in \(P_1, P_2, P_3, P_4\) and \(P_5\). In such a case we can be sure that there is always a free track in the station, where the last entered train may flow.

The inequalities (b)–(d) imply that transition \(t_{in,1}\) may fire whenever the following conditions hold: (b) Places \(P_2\) and \(P_{u,1}\) are not marked at the same time: in fact, if both these places are marked, and a new train enters the station, then the last entered train may only go to \(P_2\), thus violating safeness constraint. (c) Places \(P_3\) and \(P_{i,1}\) are not marked contemporary (similar to the previous case). (d) Places \(P_4\) and \(P_{d,1}\) are not marked at the same time (similar to the previous case).

Inequality (e) implies that transition \(t_{in,1}\) cannot fire if both places \(P_1\) and \(P_2\) are marked: the firing of \(t_{in,1}\) would allow the train in \(P_1\) to enter the up-track thus colliding with that one in \(P_2\).

Analogous considerations lead to the formulation of the other constraints in (7).

As an example, in figure 5 we have reported the monitor places relative to constraints 7.a and 7.u (\(P_{M,a}\) and \(P_{M,u}\), respectively).

Let us finally observe that, even if constraints (b)–(d) and (m)–(o) involve the firing vector, no self-loop exists between transitions \(t_{in,1}\) and \(t_{in,2}\) and the corresponding monitor places. This is due to the uncontrollability of all transitions entering places \(P_2, P_3\) and \(P_4\), and can be immediately observed by simply transforming the above constraints in accordance to the Moody's procedure.

V. Conclusions

In this paper we provided a modular representation of railway networks in terms of stations and tracks including sensors and semaphores. The resulting model is a P/T net whose transitions may be (un)controllable and/or (un)observable.

We have shown how collision avoidance constraints can be expressed as GMECs and the corresponding controller takes the form of a set of monitor places that can be computed using Moody's parametrization. This solution may lead to local deadlock: a better solution that guarantees safeness and local liveness, can be written in the form of constraints also involving the firing vector and the corresponding control structure takes the form of a monitor with self-loops.

REFERENCES


104