The Kicking Horse Pass Problem

Werner E. Kluge

Department of Computer Science
University of Kiel
D–24105 Kiel, Germany
E–mail: wk@informatik.uni-kiel.de

Abstract. *The paper addresses the organization of an orderly train traffic over a single track railway line subdivided into three sections and with sidings in between. This problem is modelled by means of two Petri nets, of which one ignores the existence of the sidings, the others includes them. Both net models guarantee essential safety and liveness properties: the track sections are by trains used in a mutually exclusive manner, deadlocks between trains moving in opposite directions are prevented, both directions are served fairly, and trains that must not compete with other trains trying to use the track (or a section of it) in the same or in the opposite direction can proceed immediately.*

1 Introduction

This paper is to discuss solutions to an organizational problem which, in a nutshell, exposes both the phenomena that typically occur in the context of concurrent activities and the measures that must be taken to coordinate these activities in some orderly form. The problem at hand - moving under some loose timing constraints several trains in both directions over a single track railway line - is of decidedly more practical relevance than, say, the famous dining philosophers problem.

This train scheduling problem has several times been used to introduce undergraduate students to basic concepts of concurrency and to system modelling with Petri nets. With little more than a very elementary notion of nets (the meaning of transitions and places, of the arcs that connect them, of the conditions under which transitions are enabled to fire, and of markings (tokens) in places) the issues that need to be dealt with can be explained in a much more comprehensible way and made more explicit than is possible with program notations and with the interleaving of program threads, as they can be found in almost all traditional textbooks on concurrency, e.g. [An91,BA90]. Not only does this textual notation require some fairly thorough understanding of state transition semantics, but it also makes some assumptions about hidden scheduling mechanisms which cannot be easily conveyed to undergraduate students who are usually not (yet) familiar with the subject.

In the sequel we will first outline the problem and then present net models for two solutions which differ from each other with respect to the assumptions that are being made about the particularities of the system configuration.

2 Problem Identification

The organizational problem to be studied here concerns the orderly coordination of trains moving in both directions over a single track railway line. This train movement must follow first principles of an orderly system behavior and also meet some loose timing constraints which render it necessary to have, at least occassionally, more than one train move along on the track. As a model for this study we choose a section of the Canadian Pacific Railway main line over the Kicking Horse pass between the station at Field in B.C. and a point called Stephen some five miles west of Lake Louise in Alberta.

This single track section, in the following referred to as the track, is sketched out in fig. 1. It is about 15 miles long, with switchyards (or sidings) at both ends to accommodate several trains waiting to move eastbound or westbound over the track. There are also two sidings along the track at points called Patridge and Cathedral. An interesting construction feature are the two spiral tunnels about halfway up the pass, in Yoho National Park, between these two points. They are to keep the gradient of the track between Stephen (at 5200 feet altitude) and Field (at 4200 feet altitude) down to about 2.2%, which is still rather considerable. Due to this steep ascent (or descent), in conjunction with the difficult passage through the tunnels (the track radius there is only about 500 feet), the maximum speed for freight trains (of up to 110 cars headed by up to 6 diesel engines) is only 20 miles / hour along the entire track, i.e., it takes about one hour for a train to go either way between Field and Stephen [Po95,CP90].

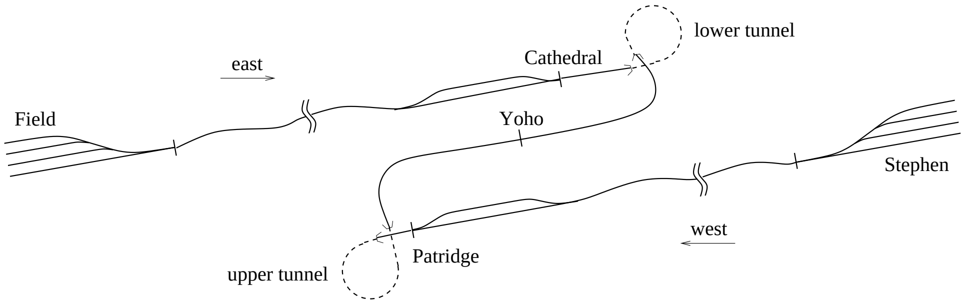

Fig. 1. Sketch of the Kicking Horse Pass section of the CP Rail main line

Current traffic load being up to 30 trains per day (15 in each direction) and increasing, some scheduling scheme must take control to handle all trains without undue delay in either direction and within the 24 hour time frame. As it must obviously be possible to have more than one train move at the same time along the track, there is a potential for deadlocks between trains going in opposite directions and, if deadlocks are ruled out by appropriate controls, for the monopolization of the track by trains going in one direction while trains trying to go in the other direction are held back for unreasonably long periods of time (or are starving) at the respective switchyards (either Field or Stephen) and beyond, possibly causing more such problems further east or west of the line [1].

In the sequel we will present two Petri net models which address this scheduling problem in different settings with respect to the availability of sidings. The first model is based on the assumption that the sidings at Patridge and Cathedral cannot be used (or are not available), whereas the second model includes these sidings to allow trains moving in opposite directions to get by each other.

3 The No-Sidings Track Model

The Petri-net models to be developed here are based on the assumptions that the track between Field and Stephen is partitioned into three sections S_1 (between

[1] Presently, train movement over the pass (and along the entire line) is mainly coordinated by radio communication with some Rail Traffic Control Center at Calgary and also among the train crews.

Field and Cathedral), S_2 (between Cathedral and Patridge) and S_3 (between Patridge and Stephen), and that at most one train is allowed to be in any of these sections at a time. A section may be entered by a train in either direction if and only if no other train is in it (it is empty).

Fig. 2 shows the basic Petri-net which just models the movement of trains, in compliance with these rules, in both directions along the track. The model for each section comprises three places, of which the one in the middle, annotated as s_i, represents the empty section, and the ones at the top and at the bottom, when marked by tokens, represent trains moving along the section eastbound and westbound, respectively. Two pairs of transitions, one for each direction, model entrance into and exit from the section. The places denoted as fd_yard and st_yard at the upper left and lower right of the net represent the switchyards at Field and Stephen, respectively. Tokens in these places represent trains waiting for entry into the track, and the single tokens in the places $s\_1 \ldots s\_3$ indicate that all sections of the track are empty.

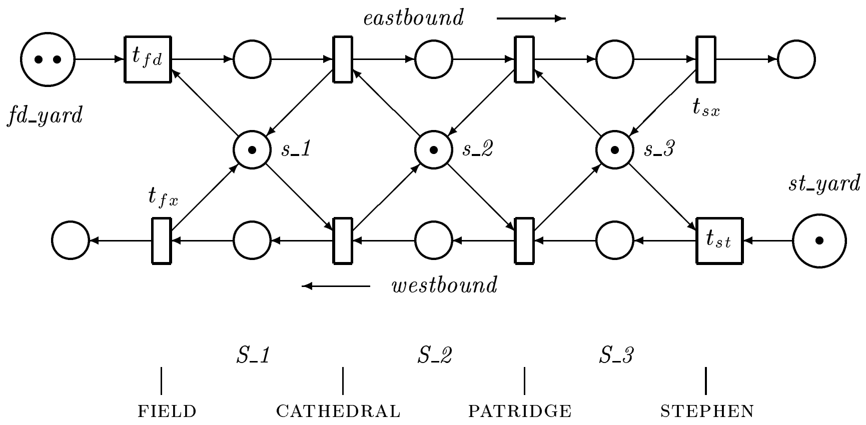

Fig. 2. Basic net model of the Kicking Horse Pass Problem

With nothing else to control their movement, a train each may enter the track from Field and from Stephen, as the respective entrance transitions t_{fd} and t_{st} both have concession to fire. The trains may proceed along the track independently of each other until eventually a deadlock as shown in fig. 3 occurs: the westbound train in section S_3 and the eastbound train in section S_2 mutually block each others entry into their next sections. This structural deadlock is characterized by the absence of tokens in the cyclic subnet formed by the two places s_2 and s_3 and by the transitions that interconnect them, with no way of ever getting tokens back into these places.

The obvious remedy to this problem is to prevent trains from entering the track in one direction while there is a train anywhere in the track going in the other direction. However, to meet the afore-mentioned timing constraints, a train moving along the track in a particular direction may be followed by more trains in the same direction as there are three sections that can accommodate them.

The necessary controls may be included in the basic track model by adding two more places *dpe* and *dpw* (for deadlock prevention east and west) which basically connect to the entrance and exit transitions at Field (t_{fd} and t_{fx}) and Stephen (t_{st} and t_{sx}), as depicted in fig. 4. Both places are initialized with three tokens each. Tokens from *dpe* may be claimed by up to three trains moving eastbound, and tokens from *dpw* may be claimed by up to three trains moving westbound.

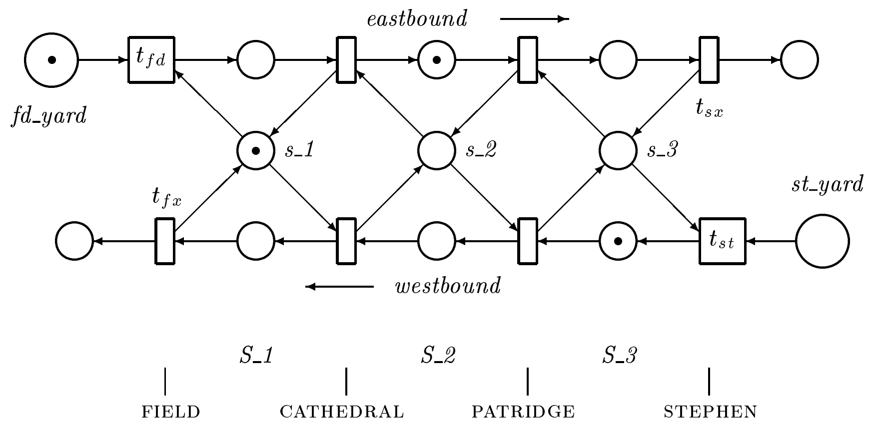

Fig. 3. Basic net model with east- and westbound trains in a deadlock situation

However, as a train entering the track from, say, Stephen to go west must also be able to withdraw three tokens from place *dpe* (which are returned upon the train moving across the section boundary at Patridge), this direction is blocked as long as there are trains in the track moving east (the same holds equivalently for the opposite direction).

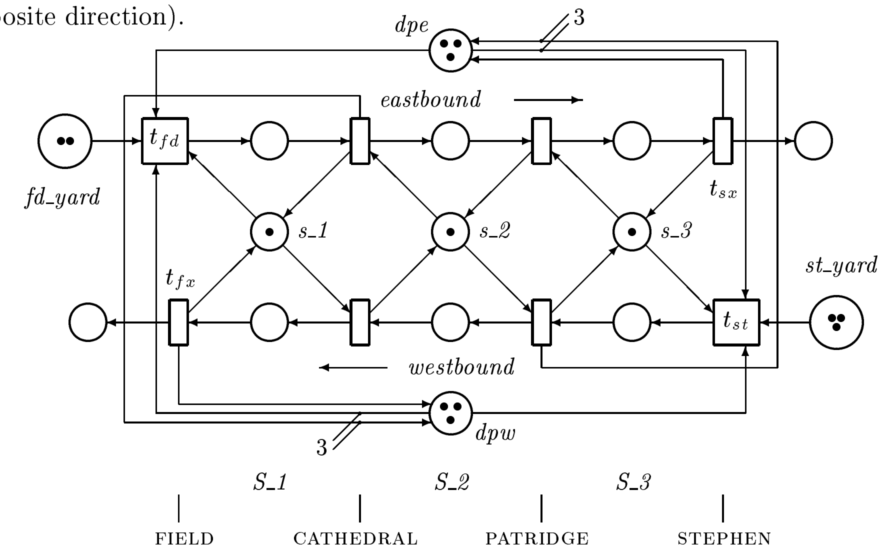

Fig. 4. The net model with deadlock prevention

Fig. 5 shows the situation of one westbound train having entered section *S_3*, with all tokens momentarily withdrawn from the place *dpe* and with one token missing from place *dpw*, which stalls all eastbound trains waiting at the Field switchyard *fd_yard*.

Fig. 6 shows two westbound trains in sections *S_1* and *S_2* of the track, with all three tokens returned to the place *dpe*. Nevertheless, all eastbound trains remain blocked as there are not enough tokens in the place *dpw* to enable the entrance transition of section *S_1*. For this to be the case again, all trains must have cleared the track and thereby fired the westbound exit transition t_{fx} which returns tokens to the place *dpw*. If it should happen that all three tokens have returned to the places *dpe* and *dpw*, i.e., no trains are in the track, but trains are waiting at both

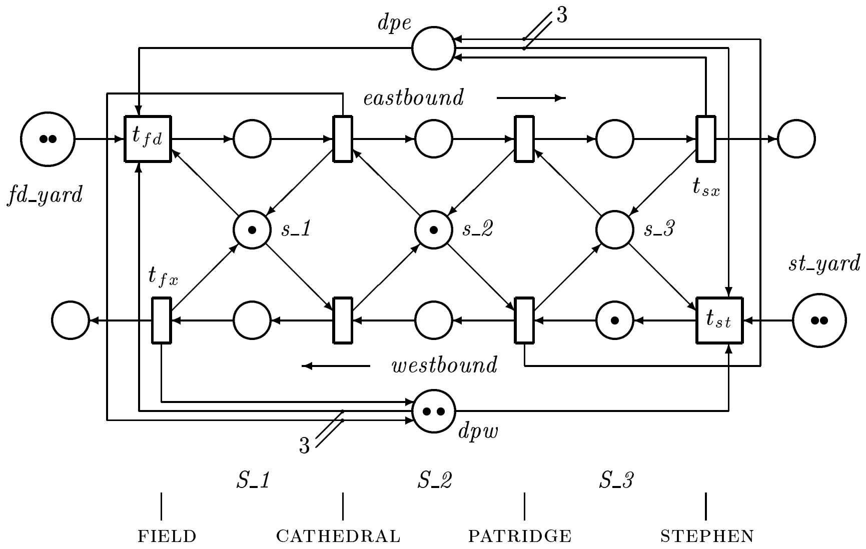

Fig. 5. Net model with deadlock prevention - one westbound train in section S_3

ends, there is still no guarantee that a change of directions is taking place. In such a situation, we have a conflict between the entrance transitions at Field (t_{fd}) and at Stephen (t_{st}) which both are enabled. Since conflicts, by their very nature, are conceptually resolved by arbitration as long as there is no other mechanism involved which renders a determinate decision, it may very well happen that further trains be granted permission to enter the track in the same direction in which the last train has left.

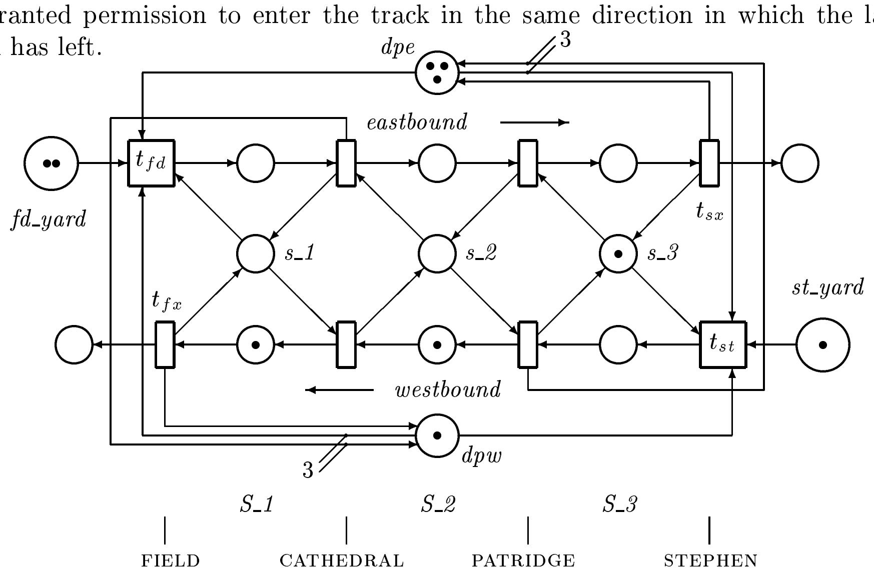

Fig. 6. Net model with deadlock prevention - two westbound trains in sections S_1 and S_2

Moreover, the added control mechanism, while successfully preventing deadlocks, potentially creates another problem. Once a train travels along the track in a particular direction, say westbound, and more westbound trains are arriving at the Stephen switchyard at intervals slightly shorter than the time it takes for one train to get from there to Field, then the track tends to get monopolized by trains going in this direction, while eastbound trains may accummulate at the Field switchyard and kept waiting for unreasonably long times. If this switchyard is getting filled to capacity by trains arriving from the west, this monopolization may even lead to deadlocks further down the single track line west of Field unless appropriate measures are taken to prevent them as well. Similar problems may, of course, occur at the opposite end of the track if it is being monopolized by eastbound trains.

This so-called starvation problem may be dealt with by a fairness regulation mechanism which, loosely speaking, puts an upper limit on the difference between the number of trains going in one direction versus the number of trains going in the other direction. As long as the train movement in both directions remains within this limit, the resolution of a conflict between an eastbound and a westbound train trying to enter the track when it is empty may be left to arbitration. However, if this limit is exhausted in favor of one direction, say eastbound, the right of way must be granted to a westbound train.

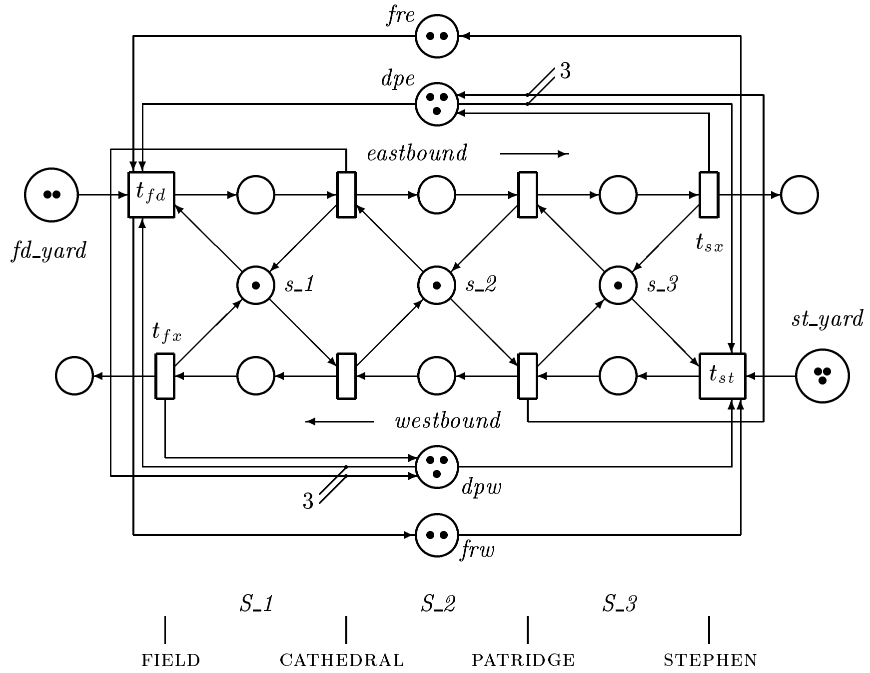

Fig. 7. Deadlock-free net model extended by fairness regulation

This fairness regulation may be added to the net model by means of another two places fre and frw which cyclically connect the entrance transitions at both ends of the track, as shown in fig. 7. An eastbound train trying to enter the track from the Field switchyard must be able to withdraw a token from place fre, and a westbound train trying to enter the track from the Stephen switchyard must be able to withdraw a token from place frw in order to succeed. A token consumed from one of these places is added to the respective other place, i.e., the number of tokens circulating about the two places remains invariant (4 in the particular case shown). With two tokens allocated to each of these places initially, a train may enter the empty track from either side (which immediately causes the deadlock prevention part to block entrance from the other side) as both entrance transitions are in conflict.

If the conflict is resolved in favor of westbound trains and two trains are moving along the track in this direction, the fairness regulation prevents further westbound trains from entering the track as the tokens in frw are then exhausted (see fig. 8). As soon as these two westbound trains have arrived at Field, the fairness regulation

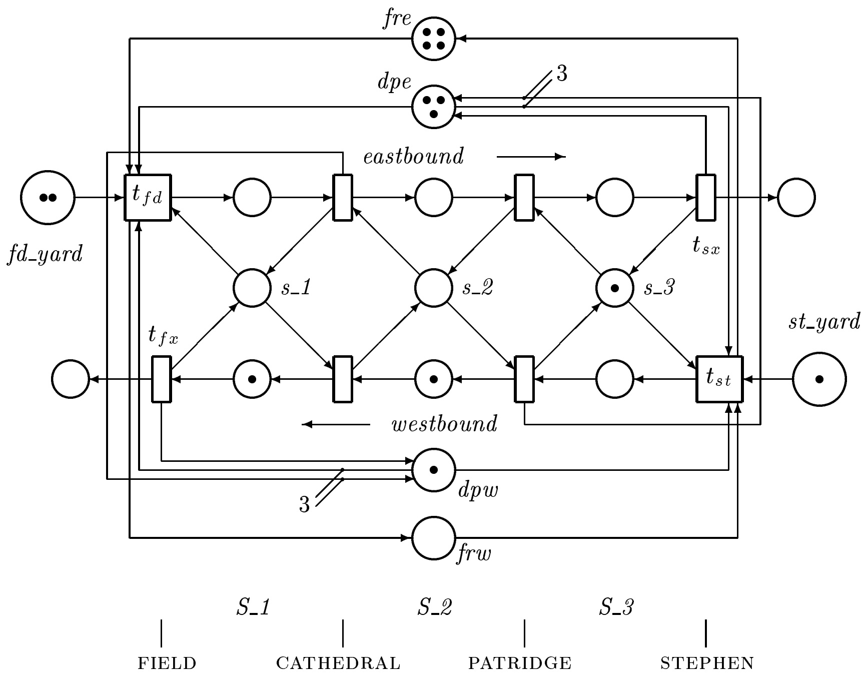

Fig. 8. Deadlock-free and fair net model - with two westbound trains in sections S_1 and S_2

leaves no other choice but to have an eastbound train enter the track from Field. This train may be followed by up to three more trains, before all tokens are being moved from fre to frw, forcing another change of directions. In between these two extremes, conflicts between eastbound and westbound trains asking for permission to enter the track when it is empty may again be resolved by arbitration. The so-called synchronic distance thus established between trains moving in both directions becomes larger with increasing numbers of tokens circulating about fre and frw. On the one hand, this increases the freedom to choose between both directions; on the other hand, provides for better utilization of the track in situations of heavy traffic: in a best case scenario, i.e., with all tokens accumulated in one of the two places, say fre, as many eastbound trains may move in succession before this direction must change. With a synchronic distance of three or (more), which allows each section of the track to be occupied by a train, and assuming that it takes about 20 minutes for a train to pass through each section (which are no more than rough approximations of real travel times), the throughput in one direction could be as high as three trains per hour. Changing directions causes a break of about one hour before the first train arrives at the other end of the track. Nevertheless, the number of trains that can be handled by this scheduling scheme may be as high as 18 trains per day and direction, based on a synchronic distance of three between eastbound and westbound trains.

Unfortunately, there is still one more problem left which, though obviously trivial, turns out to be rather difficult to include in this model. It comes about due to the fairness regulation and occurs in the extreme cases where there is no train in the track, only trains moving in one direction ask for permission to enter the track (no train is competing for entrance from the other end), but the other direction has been determined to be the one in which the next train ought to go (all tokens have accumulated in either of the places frw or fre). Unless the track is left unused until this concession is eventually taken by a train moving in the selected direction, which would waste a precious resource (the track) for no reason at all, something must be done to disable (or overrule) the fairness regulation in such situations, at least for one train.

What needs to be accomplished is to guarantee immediate entry into an empty track for a train trying to move in one direction, with no train competing for entry in the other direction. The extreme cases just described are in the net model of figs. 7 or 8 characterized, say for westbound trains, by four tokens in place fre, no tokens in frw, at least one token in st_yard, no tokens in fd_yard, and with tokens depicting the empty track in all other places.

Since Petri nets have no notion of a transition firing on the condition that one or more of its input places carry no tokens at all, we have to complement in our track model the places fd_yard and st_yard (in which tokens represent trains waiting in the switchyards at both ends of the track) by places ffd_yard and fst_yard, respectively, in which tokens represent empty sidings. Both places need to be cyclically connected with the respective entrance transitions t_{fd} and t_{st} to the track on the one hand, and with additional transitions t_{fde} and t_{ste} that model entry of trains into the switchyards from the tracks further west of Field and further east of Stephen, respectively. Now the situation in which, say, no train is waiting in Field for an eastbound passage to Stephen can be represented by as many tokens in place ffd_yard as there are sidings in the switchyard. Every train arriving from the west claims one of the sidings by removing a token from ffd_yard and putting it into fd_yard.

Fig. 9 shows just those parts at both ends of the full track model of figs. 7 and 8 where these extensions, together with the mechanism that overrules the fairness regulation, are being included. The crucial part in this net is played by the two transitions t_w and t_e connected to the places fre, frw on the one hand, and to the places that represent the switchyards on the other hand. In the situation shown in fig. 9 we have an empty switchyard at Field (assuming a total of $c_{fd} = 4$ sidings) as all tokens are in place ffd_yuard. The fairness regulation has all tokens accumulated in place fre, trying to enforce access to the track for an eastbound train which is not there, while two trains are waiting at the Stephen switchyard to go west.

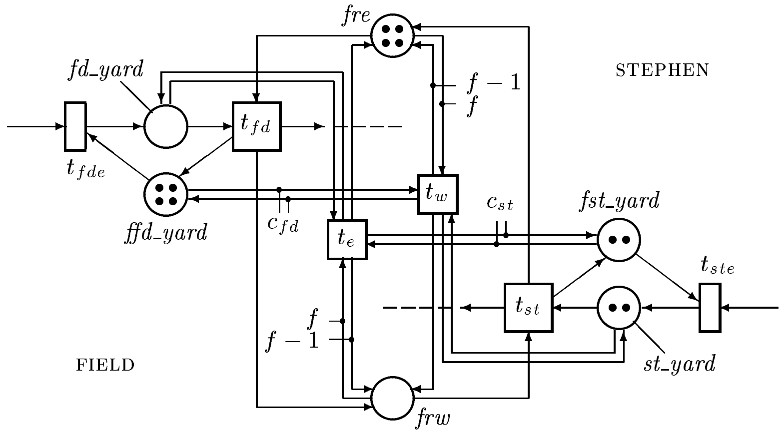

Fig. 9. Net fragment which enables immediate entry into an empty track for a train moving in one direction (westbound in the situation shown) while no train competes for entry from the other direction

Now, the transition t_w enters the game: it is connected to the place ffd_yuard by two arcs of weights c_{fd} pointing in opposite direction and to place st_yard by another two arcs of weights 1, again, pointing in opposite directions. Two more arcs connect t_w to the place fre, the one pointing towards t_w has weight f, the one pointing towards fre has weight $f-1$, with f denoting the total number of tokens that is circulating about fre and frw. Yet another arc of weight 1 leads from t_w to the place frw.

Thus, the transition t_w is enabled to fire if there are f (all) tokens in fre, c_{fd} (all) tokens in ffd_yuard, and at least one token is in st_yard, i.e., it tests for exactly the conditions under which the fairness regulation must be overruled. The firing of this transition moves one token from fre to frw but leaves intact the number of tokens in the other two places to which it is connected bidirectionally (i.e., the conditions represented by these places are, in fact, side conditions). As a result, the entrance transition t_{st} for westbound trains in now enabled for the passage of exactly one train. Further trains may follow in this direction by repeatedly enacting the same mechanism as long as no eastbound trains arrive at the Field switchyard. As soon as this becomes the case, the fairness regulation takes over again, i.e., the next train is going east after all westbound trains have left the track.

The full Petri-net model which we have arrived at now guarantees that

– the three sections of the track may be used by trains only in a mutually exclusive manner, i.e., by at most one train at a time;
– no two trains get into a deadlock situation: if they demand entry into the tarck from opposite directions, only one of them can proceed;
– trains are not unduly delayed: entry into the track is by a fairness regulation mechanism granted within some finite synchronic distance to all trains that apply (compete) for it;
– a train requesting entry into the track which has no competitor demanding entry from the other side gets permission to proceed immediately;

i.e., the net models first principles of an orderly system beavior in that it meets essential safety and liveness properties.

4 The Track Model with Sidings

We consider now a modified track model which includes the sidings at Patridge and Cathedral where trains going in opposite directions may get by each other. In situations of dense traffic these sidings may help to move more trains between Stephen and Field than without them.

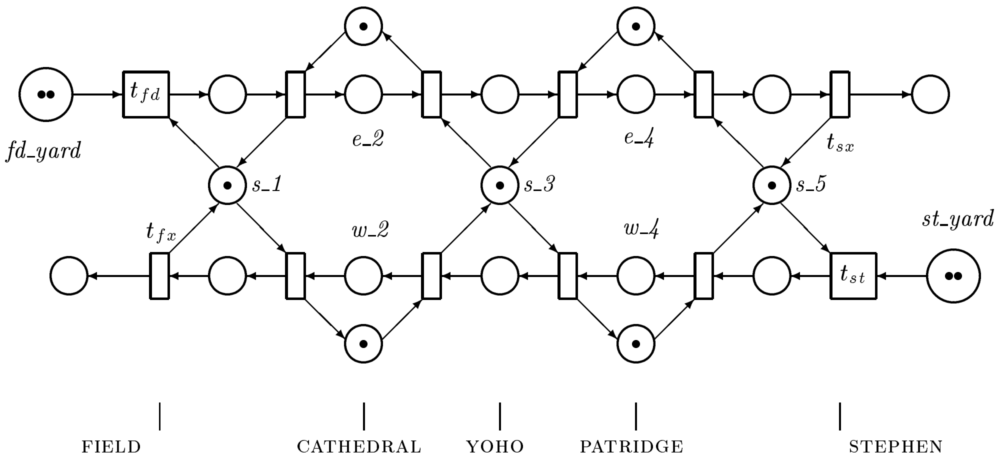

Fig. 10. Basic net model of the Kicking Horse Pass problem with sidings included

Fig. 10 shows the basic net model of the track. It includes the three single track sections now labeled s_1 (between Field and Cathedral), s_3 (between Cathedral and Patridge) and s_5 (between Patridge and Stephen). The sidings e_2 and e_4 at Cathedral and Patridge, respectively, may only be used by eastbound trains, whereas the sidings w_2 and w_4 at these points may only be used by westbound trains. The figure shows the initial state in which all sections and all sidings are empty, and two trains each are waiting at the Field and Stephen switchyards for entry into the track.

If trains start moving from both ends of the track under no other constraints than mutually exclusive entry into sections and sidings, they may again end up in a deadlock as depicted in fig. 11: the two trains that have made it to the sidings at Cathedral cannot proceed any further as the sections they need to enter next are already occupied by trains trying to move in the respective opposite directions (an equivalent deadlock could occur at the Patridge sidings).

This deadlock is characterized by the cyclic subnet made up from the places s_1, s_2 and from the places complementary to e_2 and w_2 of which none carries a token. As a consequence, none of the entry and exit transitions of the sidings e_2 and w_2 has concession and will ever get it since the transitions that could inject tokens into the places s_1 or s_3 from somewhere else also belong to cyclic subnets which are left without tokens.

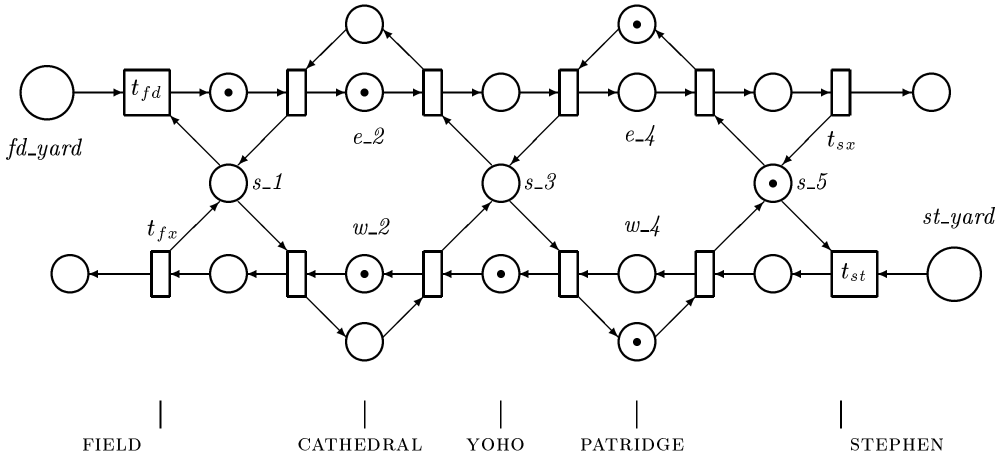

Fig. 11. A deadlock situation in the net model of the Kicking Horse Pass Problem with sidings

The deadlock in fig. 11 can obviously be prevented if at least one of the track positions involved can, by some additional controls, be kept unoccupied at all times, which is equivalent to having in all cyclic subnets at least one place marked by a token under all possible token distributions. For instance, if

- one of the eastbound trains is held back at the Field switchyard, leaving the section s_1 empty, both westbound trains could pass through, before the eastbound train at the Cathedral siding e_2 can move on;
- the second of the westbound trains is held back at the Patridge siding w_4, then both eastbound trains could pass through and, as soon as the second eastbound train has reached the Cathedral siding e_2, the westbound train waiting at w_2 could move on as well;

– the two westbound trains are in positions s_3 and w_4, and w_2 is empty, then the train in s_3 can proceed to w_2, which brings about the same situation as under the second item.

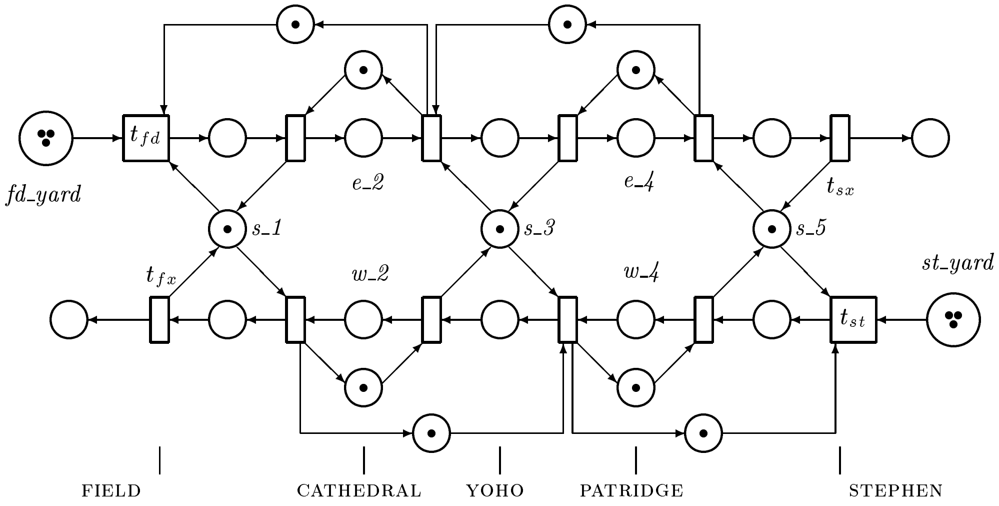

Fig. 12. Conservative deadlock prevention in the net model of the Kicking Horse Pass with sidings

A simple conservative solution to this deadlock problem consists in letting at most one eastbound train each enter section s_1 or siding e_2 and section s_3 or siding e_4. Likewise, at most one westbound train each is permitted to enter section s_5 or siding w_4 and section s_3 or siding w_2. The extensions necessary to include these measures into the net model are shown in fig. 12. The token distribution again depicts the initial state of an empty track, with three trains waiting on each side for eastbound and westbound passages.

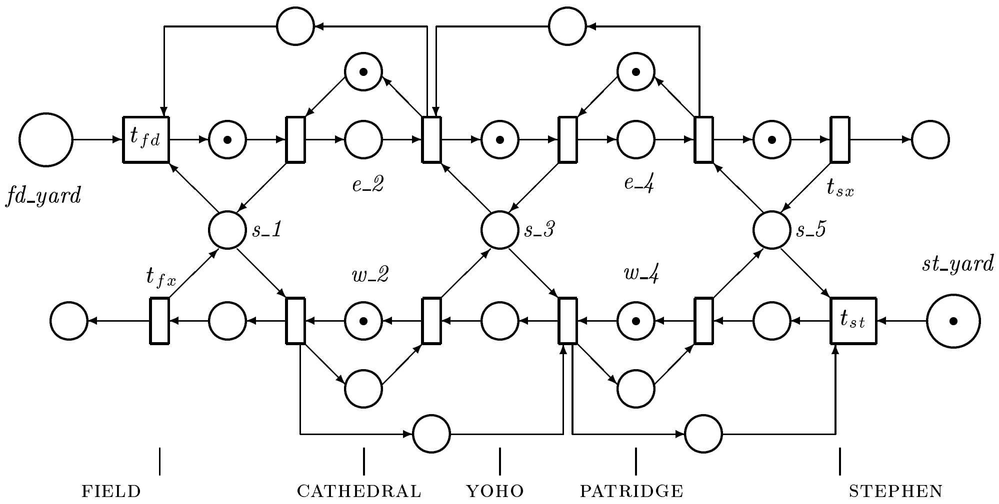

Fig. 13. Two westbound and two eastbound trains in the track under conservative deadlock prevention in the net model with sidings included

These additional controls restrict to one each the number of westbound and eastbound trains that either approach or are in one of the sidings at Patridge or Cathedral, i.e. it is ensured that as soon as both trains have reached these sidings, they may also proceed beyond. A typical situation of this kind is depicted in fig. 13 where we have eastbound trains in the sections s_1, s_3 and s_5, and westbound trains in the sidings w_2 and w_4. The eastbound trains are free to move on to the sidings e_2 and e_4, and the train in s_5 may leave the track at Stephen, whereupon the westbound train at w_2 gets permission to pass through section s_1 and leave the track at Field. Likewise, the eastbound train that arrives at the siding e_4 gets immediate permission to pass through section s_4 and leave the track at Stephen. The conflict for access to the section s_3 between the westbound train left at w_4 and the eastbound train left at e_2 may be decided either way as it merely affects the order in which they may proceed. Both trains can leave the track eventually even if further trains are entering from the respective opposite directions.

This deadlock prevention mechanism permits at most five trains to be in the track at the same time, with three trains going in one direction and two trains going in the other direction. Of course, at most three of these trains can be actually in motion along the sections s_1, s_3 and s_5, whereas the other two must wait in the sidings.

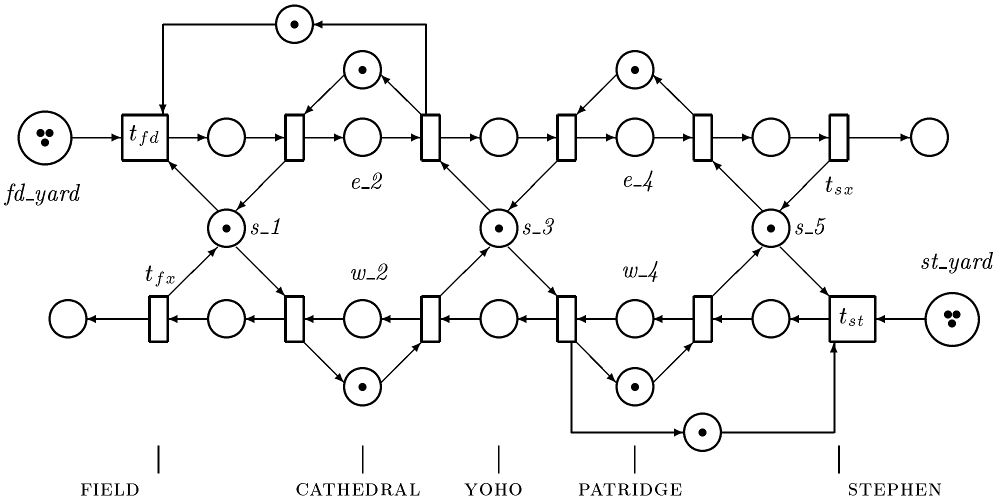

Fig. 14. Minimal deadlock prevention in the net model of the Kicking Horse Pass with sidings

However, for a minimal solution of the deadlock prevention problem it suffices to restrict to at most one eastbound train the usage of section s_1 or siding e_2 at Cathedral and to at most one westbound train the usage of section s_5 or siding w_4 at Patridge, and to leave otherwise the usage of sections by eastbound or westbound trains to arbitration. This solution, of which the net model is shown in fig. 14, perfectly complies with the afore-mentioned criterion that at most three trains may either approach or be in the sidings at Patridge or Cathedral. It requires only two of the four feedback loops that had to be introduced in the conservative deadlock prevention model of figs. 12 and 13.

Fig. 15 shows a possible situation under this relaxed deadlock prevention mechanism with six trains in the track, of which three each are eastbound and westbound. Again, only three of them - those in sections s_1, s_3 and s_5 - may be in motion.

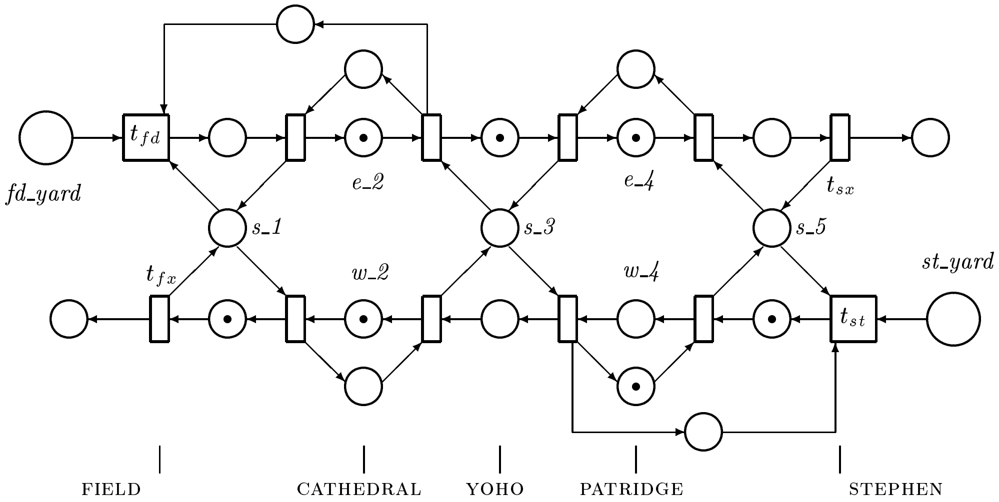

Fig. 15. Six trains in the track under minimal deadlock prevention

Another two trains each are waiting at the Field and Stephen switchyards for eastbound and westbound passages.

Under such dense traffic in both directions the train movement can now be organized such that for each train leaving a particular section and entering a siding (or leaving the entire track), another train waiting in the complementary siding (or in the respective switchyard) may enter the section in opposite direction, i.e., each section may alternatingly be used by an eastbound and by a westbound train. Assuming a traveling time through a section of about 20 minutes and a negligible waiting time at the sidings if the arrivals of east- and westbound trains are nearly perfectly synchronized, a train each may arrive and depart in intervals of 40 minutes at both ends of the track. This amounts to a throughput of 24 trains per day and direction at the maximum, or to a 50% improvement over what can be accomplished with the mode of operation described in the preceding section, which does without the sidings at Patridge and Cathedral.

Beginning with the particular situation of fig. 15, trains could proceed alternatingly as follows: as soon as the train going west in section s_5 arrives at the siding w_4, the train waiting in siding e_4 may move on east through section s_5. This enables the train that moves east in section s_3 to enter siding e_4, freeing s_3 for the train that now waits in w_4. As soon as the eastbound train now in section s_5 leaves the track at Stephen, another westbound train may enter from there. Likewise, after the westbound train in section s_1 has left the track at Field, and the eastbound train waiting in siding e_2 has, in the meantime, advanced to section s_3, another eastbound train may enter the track at Field and advance to siding e_2, before the westbound train waiting in w_2 may follow.

The alternate usage of sections by east- and westbound trains may be enforced by degenerate fairness regulation cycles layed over their entrance transitions which have only one token each circulating about their two complementary places. These alternators too must be augmented by subnets which overrule their decisions in cases where trains are waiting for passages in the directions opposite to those actually chosen, with no trains competing for passages in the chosen directions.

A fairness regulation which, as in the no-sidings track model of the preceding section, mediates over the entire track between eastbound and westbound trains is superfluous here as the local alternators prevent its monopolization by trains going in one direction.

References

[An91] Andrews, G.R..: *Concurrent Programming, Principles and Practices* , Benjamin / Cummings Publishing Company 1991

[BA90] Ben-Ari, M.: *Principles of Concurrent and Distributed Programming*, Prentice Hall International Series in Computer Science 1990

[GeLa78] Genrich, H.J.; Lautenbach, K.: *The Analysis of Distributed Systems by Means of Predicate/Transition-Nets*, in: Semantics of Concurrent Computation, Lecture Notes in Computer Science, No. 70, Springer, 1979, pp. 123–146

[Jen90] Jensen, K.: *Coloured Petri Nets: A High-Level Language For System Design and Analysis*, in: Rozenberg, G. (Ed.): Advances in Petri Nets 1990, Lecture Notes in Computer Science, No. 483, Springer 1991, pp. 342–416

[Po95] Pole, G.: *The Spiral Tunnels and the Big Hill*, Altitude Publishing Canada Ltd. 1995

[CP90] *CP Rail HeavyHaul System Alberta, Calgary, Revelstoke and Vancouver Division, Time Table 85*, CP Rail Internal Document 1990

[Rei85] Reisig, W.: *Petri Nets*, EATCS Monographs on Theoretical Computer Science, Vol. 4, Springer, 1985