A Generalized Stochastic Petri Net Model for Performance Analysis of Trackside Infrastructure in Railway Station Areas under Uncertainty*

Malte Schmidt1, Norman Weik2, Stephan Zieger2, Anke Schmeink1 and Nils Nießen2

Abstract—The paper presents a stochastic, infrastructure centered modeling approach for capacity analysis applications in long term planning of railway stations. A formal modeling approach of train operations using generalized stochastic Petri nets is proposed. The model is analyzed based on the embedded Markov chain, which allows for quality based performance metrics including utilization, blocking probabilities and waiting times. The model is validated and tested against simulation in an application scenario for a medium size railway station. In addition, model extensions to large-scale systems are discussed.

I. INTRODUCTION

Complicated approval procedures and long asset lifetimes require intensive planning in railway projects. Capacity analysis, which deals with assessing the traffic load that can be sustained by the infrastructure at a high level of quality, is a vital planning step in this context. Whereas significant efforts have been undertaken to analyze and standardize capacity evaluation of railway lines [1], [33], [36], such that lines are well-understood, the situation is different for railway nodes.

On the tactical level, traffic planning and optimization in railway station areas have successfully been tackled in the station routing [5], [24], [26] and the train platforming problem [9], [10], [32]. Recently, robust scheduling techniques [4], [7], [12] have found widespread application in this area.

As these approaches require information on train arrival and departure times, they are unsuited for strategic infrastructure dimensioning where timetable information is scarce and infrastructure capability regardless of a specific timetable is to be assessed. This is why analytic and stochastic models, as well as simulation studies are prevailing in long term station performance analysis. UIC Code 406 [36] extends the schedule compression method, which is based on infrastructure utilization, to track groups and station threads. In probabilistic methods compiled and analyzed in [27], station capacity is assessed in terms of blocking probability and waiting times at station entry and exit. Similar techniques are applied in [25], where additionally the interface to the UIC compression method is explored.

Decomposition techniques for bottleneck analysis of station areas [31] have frequently been used [22]. To assess entire route nodes, an aggregation of infrastructure segments – so called serial route nodes – has been performed [29]. To cope with uncertainty in traffic demand, approaches involving discrete event simulation [14], [15] are widespread. Whereas the previously mentioned stochastic approaches suffer from a lack of exactness in representing train operations due to aggregation or decomposition, simulations are time consuming given heavy-tailed train processing times [3].

Petri nets, which allow for a detailed representation of operating conditions in station areas and dependencies imposed by the safety system, provide a means to mitigate these problems. By allowing for an exact representation of the train control logic, while providing flexibility in transition modeling, they are particularly suited for long-term planning purposes with uncertainty. In [20], [28] and [34], Petri nets are used to represent railway control systems and interlockings focusing on the safety aspect. In traffic planning, colored Petri nets have been used as a formal abstraction of timetable generation as early as 1996 [30]. A detailed control model for railway operations based on the interlocking logic and its application to deadlock prevention is investigated in [18]. Ref. [21] presents a fully automated approach to set up Petri net simulation models for railway networks from elementary railroad infrastructure data and its application to the Oslo subway network.

At a high-level train operation perspective, Petri nets have been used to describe the train processing on lines [16], [38] and in stations [2], [35]. Delay propagation in stations is discussed in [40] and has been picked up in [17] and [39]. Still, the use of formal methods with the goal of station capacity analysis is limited: In Ref. [37], an interval timed Petri net for traffic modeling in railway stations is presented and hard performance bounds for a single operation hour are derived. However, to the best of our knowledge, Burkolter’s work [8] remains the only approach to perform formal Petri net based capacity analysis for extended time frames and station areas. A two-level approach is used: On the aggregated base level, Petri nets are used to create a tentative timetable based on a macroscopic train event graph including interactions between trains, but disregarding microscopic route exclusions in stations. This event-activity graph is then evaluated using max-plus techniques. On the second level, this tentative timetable is checked for microscopic feasibility in station threads. If necessary, an iterative adjustment to the train sequence in the base layer is made. Capacity is assessed based on the critical path in the max-plus setting, i.e. the densest possible train sequence that can be generated for given train service requests (also see [19]).
In this paper, we present a generalized stochastic Petri net (GSPN) model for performance modeling of railway station areas related to [8] and [37] and demonstrate how the framework can be used for station capacity analysis. The focus is on long term strategic infrastructure and line planning with uncertain traffic demand, where no detailed timetabled concept and only a rough operating concept is available. Our model goes beyond stochastic approaches prevalent in this area by:

- accounting for complex nonlinear interactions between trains in the Petri net. Asynchronous track occupation and release is explicitly considered. Such interactions are not incorporated in existing models, where a decomposition of station areas into track groups and route nodes is widespread [36],
- extending the performance analysis to more detailed quality-oriented capacity metrics such as the distribution of system times or blocking probabilities at station threads and at platforms. Corresponding performance indicators derive from a state space based analysis of the embedded Markov chain.

The paper is organized as follows: In Section II we introduce some elementary notions of train operations in railway station areas. The GSPN station model, as well as a formalized approach to identify and include partially exclusive shared resources in route nodes and station areas, is discussed in Section III. Section IV presents an analysis of the model’s performance for a medium size railway station.

II. TRAIN OPERATION IN STATION AREAS

Railway Station Areas (RSA) consist of platform tracks and switch areas. Train processing tasks such as passenger exchange and goods transfer are performed at platforms, whereas switch areas connect platforms to adjacent railway lines. For switch areas, a subdivision into serial route nodes, i.e. connected areas that can only be allocated to one train at a time has proved effective [31]. In the context of this work, serial route nodes and platform tracks are referred to as shared resources. A schematic representation of a railway station including the shared resources is depicted in Fig. 1.

Train operation in station areas is based on train paths defined in the interlocking system. Running trains requires the allocation of a train path upon station entry and exit, which involves the simultaneous blocking of all shared resources involved in the train path. To describe route dependencies, formalized route exclusion tables that describe compatibility of routes in interlocking are widely used [27].

III. GSPN MODEL FOR PERFORMANCE MODELING OF RAILWAY STATION AREAS

A. Generalized Stochastic Petri Nets

Generalized Stochastic Petri Nets (GSPN) are an extension of Petri Nets (PN) incorporating two types of transitions: immediate transitions and timed transition. Immediate transitions correspond to transitions in classic Petri Nets and fire immediately if enabled. Timed transitions, by contrast, fire after an exponentially distributed time \( t \sim \text{Exp}(\lambda) \). Immediate transitions have priority over timed transitions. In case multiple immediate transitions are enabled, firing order is according to a specific firing policy.

The two types of transitions yield vanishing and tangible markings in GSPNs. A vanishing marking is a marking in which at least one immediate transition is enabled. The holding time in vanishing states is always zero. Tangible markings are markings bearing no immediate transitions. Here, the exit rate is the sum of the transition rates of all enabled transitions.

GSPNs also allow for arc multiplicity and inhibitor arcs. Inhibitor arcs connect transitions and places and disable transitions in case the number of tokens in the place connected to the transition exceeds the multiplicity of the inhibitor arc.

B. GSPN Modeling of Train Traffic in Railway Station Areas

Following up on similar practices in delay propagation modeling for railway systems [6], [8], [19], different train patterns (in terms of routing / stopping policy / speed) are modeled as individual train lines, i.e. successions of train events and activities. For a single train line this yields the model layout in Fig. 2.

The train arrival process to the station area is modeled by the timed transition \( t_{\text{arrival}} \) and an immediate transition \( t_{\text{enter}} \), which is enabled whenever the platform and serial route nodes required by the entry route are available. In terms of the GSPN, this relates to a marking with tokens in the places corresponding to shared resources (\( p_R \) in the simplistic representation in Fig. 2). \( t_{R_1} \) and \( t_{R_2} \) are timed transitions describing the train running time upon station entry and exit. \( p_{\text{dwell}} \) denotes the halting place at the platform, which is connected with a timed transition \( t_{\text{dwell}} \) modeling train stopping time. Once the stopping time has passed, the train is ready to depart by an immediate transition \( t_{\text{delay}} \) in case the exit route (route 2) is available. In this case, a token is transferred to \( p_{R_1} \), thus enabling entry routes to the platform.
Joint use of the infrastructure imposes dependencies between train lines and connections in the GSPN graph. Figure 3 depicts the GSPN model for a simple crossing station and two train lines. Train line 1 arrives from A and proceeds to B via track 1, train line 2 arrives from B and proceeds to A via track 2. Shared resources $X_1$ and $X_2$ correspond to the serial route nodes to the left and right of the two platforms.

Rail operation requires simultaneous utilization of multiple resources as the allocation and release of track segments is not performed synchronously. A track segment can only be released for other trains once the departure of the complete train has been detected by track clearance sensors (axle counters or track circuits). Stochastic network models often neglect this fact, hence underestimating track occupation times. The GSPN model allows to incorporate asynchronous track occupation and release by including a timed transition ($t_{release}$) in the feedback loop after running time on route $R_1$ and station dwell time have been completed (see Fig. 4).

Fig. 4: GSPN model with separate route and platform release.

C. Performance Evaluation

For performance evaluation the GSPN model is transformed to a continuous time Markov chain (CTMC). Vanishing markings are eliminated from the GSPN’s reachability graph as multiple simultaneous state space transitions cannot be modeled by CTMCs. In this work, vanishing markings are eliminated in a post-elimination step according to the approach proposed in [11]. While some details on train interdependency are lost in this approach, the soundness of the effective statistical behavior of the system is maintained.

Different performance measures for capacity analysis of station areas derive from the generator matrix of the underlying CTMC. Arguably, the most widespread performance metric in railway capacity analysis is the utilization of infrastructure components. By solving the balance equations, utilization ratios for platform tracks and shared resources can be obtained from the stationary probabilities that tokens are in the corresponding place. Other metrics used in the railway context are the waiting and delay probability of trains upon station entry or exit. These quantities directly derive from the stationary probability of the markings of $p_{queue}$ or $p_{delay}$.

Mean sojourn times of trains in the station area can equally be calculated from the means of the station holding times in the CTMC. In addition to these metrics, the state-space representation of the GSPN yields full information on the distribution of sojourn times and delays (see [13]), such that non-compliance with a pre-defined level of service widely used as a capacity metric in railways [1] can be studied. In addition, transient analysis, e.g. in case of disturbances or changing traffic load during peak hours, is also possible.

Unlike previous probabilistic methods for railway stations, GSPN modeling of train operations in station areas has the advantage that nonlinear effects resulting from complex interactions between trains in station areas are included in the model in a natural way. In particular, blocking in the left station thread may affect the right station thread due to prolonged track occupation times at the platform tracks or starvation of succeeding train arrivals.

IV. APPLICATION OF THE GSPN MODEL IN STATION PERFORMANCE ANALYSIS

In the following, we analyze the performance of the GSPN approach by comparison to simulation. In the first step, the model is validated by comparison to discrete event simulation of rail traffic using the same distributional assumptions as in the GSPN model. In a second step, the model performance is assessed by comparison to a more detailed simulation tool, which includes physical calculation of train running times.

A. Application Scenario

For the analysis, a station layout and traffic concept motivated by the former Frattamaggiore station previously discussed in [27], is used. It consists of 4 platform tracks, of which 3 are directional, and 1 is bidirectional (see Fig. 5). Route and platform length are assumed to be 700 m and 500 m. Traffic concept, routing and train characteristics used in the case study are depicted in Tab. I.
TABLE I: Traffic parameters in the case study.

<table>
<thead>
<tr>
<th>Train length</th>
<th>Trains/24 hours</th>
<th>Dwell time</th>
<th>Accel.</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{in}$-1-$B_{out}$</td>
<td>300</td>
<td>56</td>
<td>2.0</td>
<td>0.15</td>
</tr>
<tr>
<td>$A_{in}$-2-$B_{out}$</td>
<td>300</td>
<td>56</td>
<td>2.0</td>
<td>0.15</td>
</tr>
<tr>
<td>$A_{in}$-4-$B_{out}$</td>
<td>100</td>
<td>8</td>
<td>1.5</td>
<td>0.20</td>
</tr>
<tr>
<td>$B_{in}$-3-$A_{out}$</td>
<td>300</td>
<td>112</td>
<td>2.0</td>
<td>0.15</td>
</tr>
<tr>
<td>$B_{in}$-4-$A_{out}$</td>
<td>100</td>
<td>8</td>
<td>1.5</td>
<td>0.20</td>
</tr>
</tbody>
</table>

An illustration of the GSPN model for the station is depicted in Fig. 6. The size of the waiting space at the adjacent lines is restricted to 3 to reduce state space size in the CTMC representation. In the GSPN, this is modeled by an inhibitor arc with multiplicity 3 in the arrival process. Asynchronous track release is not depicted here.

B. Validation of the GSPN model

For validation, stationary asset utilization, mean sojourn times, and mean waiting times are compared to a probabilistic simulation of train operations in railway stations. Timed transitions are governed by exponential distributions both in the GSPN and the simulation setting. Simulation results have been averaged over 100 runs with a total simulation time of $10^6$ minutes of train operation for each run to allow for comparison with stationary GSPN results. Asynchronous track release is not considered in the validation step.

The results are depicted in Fig. 7. Fig. 7a provides an infrastructure centered view on the utilization of the shared components. The mean station sojourn times of trains on the five corresponding routes and the mean waiting times at the station borders are depicted in Figs. 7b and 7c. Finally, the overflow probability, i.e. the probability that an arriving train finds the waiting space (of size 3) at the station border full and is lost, is depicted in Fig. 7d.
It can be seen that the agreement between GSPN results and stochastic simulation is almost perfect; relative errors are below 0.8% for all values. The overflow probability is 0.57% for station thread B and 0.16% for station thread A. While the values are non-negligible, the error made by cutting the waiting space at size 3 is not likely to seriously affect the results. The significantly higher value for rightbound trains can likely be explained by the fact that rightbound trains are evenly spread over platforms 1 and 2, such that the platform area approximately acts as a double channel service unit, whereas almost all leftbound trains are routed via Platform 3, and hence effectively experience a single channel system.

C. Comparison to simulation results

To assess the exactness of the GSPN model, results are compared to a more realistic simulation of train operations in station areas, which includes a running time calculation module. Train driving characteristics used for the simulation are depicted in Tab. I. GSPN running time parameters are calibrated to the average runtimes on the corresponding infrastructure segments rather than the maximum speed. Train routes \( A_{in} - 4 - B_{out} \) and \( B_{in} - 4 - A_{out} \) are not considered to reduce state space size when asynchronous track release is considered. Again, the simulation results are averaged over 100 runs with \( 10^8 \) simulation minutes of train traffic.

Results are depicted in Fig. 8. It can be seen in Fig. 8a that the GSPN approach gives an adequate approximation of the simulated utilizations. Whereas running times during station entry and exit are almost exact, dwell times at station platforms are slightly overapproximated. This can probably be explained by the fact that our GSPN approach overestimates variation of processing times by assuming exponential holding times in timed transitions. This particularly seems to affect the blocking probabilities at station exit and entry and hence manifests in increased platform occupation times and prolonged station entry waiting times (visible in Fig. 8c). Overall sojourn times in the station area seem less affected.

D. Limitations and possible extensions

An even better representation of blocking effects could possibly be achieved by extending the GSPN model to phase-type distributed firing times. However, this would yield a further increase of the state space. In Tab. II, basic model parameters and solution times for the GSPN approach both with and without asynchronous track release are given.

It can be seen that the model performance notably depends on the number of train lines. By increasing from 3 to 5 train lines – typical values for small and medium size station upon suitable aggregation – the state space grows from 6446 to 289071 states. Yet, large stations easily involve 10 or more different train lines that cannot be aggregated. Still, we find that the size of the state space is not the primary limitation of our approach. Sparse CTMCs with several millions of states have been shown to be tractable in probabilistic model checking [23]. The main effort currently stems from matrix inversions in the elimination of vanishing markings in the transformation of the GSPN to a CTMC based on the procedure described in [11]. Here, the main problem is not the number of operations – which are governed by the degree of sparsity (occ. < 0.1%), but fill-in in the inversion of the routing matrix between vanishing markings. By performing the elimination “on the fly” during model setup (also see [11]), we are confident that a significant gain in performance also making large stations accessible can be achieved. In addition, state space reduction techniques as, e.g. described in [37], can be used to improve performance in case of large-scale systems.

![Bar chart](image-url)

**Fig. 8:** Comparison of GSPN results to a realistic simulation with physical calculation of train running times. Simulation results averaged over 100 runs with \( 10^8 \) simulation minutes.

**TABLE II:** Computational performance of the GSPN model.

<table>
<thead>
<tr>
<th>Modeling Technique</th>
<th>Without Separate Release</th>
<th>With Separate Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train Lines 3</td>
<td>0.2min 6446</td>
<td>7min 2326</td>
</tr>
<tr>
<td>Train Lines 5</td>
<td>75min 289071</td>
<td>exceeding memory</td>
</tr>
</tbody>
</table>
V. CONCLUSIONS

We have presented a new GSPN modeling approach for the description of train operations in railway station areas in view of uncertain traffic demand. The approach is particularly suited for strategic capacity analysis tasks typically encountered in infrastructure dimensioning and line planning. By adopting GSPNs for capacity planning, a formalization of capacity modeling of station areas that incorporates nonlinear blocking effects not included in analytic, probabilistic and queuing based approaches used to date, has been achieved. In addition, our solution based on the embedded Markov chain makes more detailed capacity measures such as the marginal waiting times at specific network positions accessible. Information on distributional performance characteristics or transient system behavior can also be obtained, such that the approach can also be applied to analyze the effects of disturbances in the future.

REFERENCES