Article for the hEART 2014 conference

Productivity and Scheduling of Intermodal Terminals in a Railroad Network Context

René Schönemann*

August 15, 2014

This paper investigates the productivity of intermodal transhipment terminals under the consideration of influences from the railway network. Whereas the maximum turnover of a terminal is usually stated in loading units handled per year, its real productivity depends on external influences such as the train timetable, the service quality of the railway network it is connected to, or internal process flows. Using event-based simulation methods, the study shows a network-oriented rail yard scheduling approach that investigates the transport flow in intermodal transshipment terminals.

1 Motivation

The transshipment capacity of intermodal hubs is usually measured separated from influences of the railroad network using terminal simulation tools, cf. e.g. (Carteni and De Luca, 2009). Also railway yard operations are usually simulated in isolation using specific microscopic software tools,

*Technische Universität Berlin, Chair of Railway Track and Operations, RSchoenemann@railways.tu-berlin.de

cf. e.g. Nash and Hürlimann (2004). To the best of my knowledge, a combination of both, logistics and railroad simulation methods that evaluate the productivity of intermodal hubs have not been studied until today on a microscopic level. The analysis presented here aims to combine both. It shows how the integrated analysis can optimise the hub as a whole and identify overall bottlenecks.

By studying the operations of several European freight hubs it has been found that there are numerous factors that are limiting their theoretically calculated capacity when loading to and from rail. A preliminary analysis of railway-based process and operation flows came to the conclusion that processes between the intermediaries of a transport chains are poorly adjusted.

Consequently, the productivity of the rail yard and therewith the productivity of the whole transshipment hub decreases. Uncoordinated transport flows and thus disruptions and delays in cargo forwarding have a negative effect on the transport times and on the productivity of railway facilities in hubs (long dwell times for trains, tracks blocked with unproductive processes). The identified problems are eminent for larger, highly utilised hubs such as container terminals in seaports.

1.1 Transport and Process Flows in Logistics Hubs

Logistics hubs are nodes in a network where traffic flows merge and transport chains, means of transport and markets are connected. Logistics hubs are also the locations where transport flows break. High efforts for coordination and many involved actors lower the theoretically calculated capacity when loading to and from rail. For the purpose of this paper, a logistics hub is understood as a conjunction of at least one transport system with a transshipment system as illustrated in Figure 1.

In detail it comprises:

- Railway node: For example a marshalling yard where railway operations such as shunting is carried out
- Transshipment System: Loading and unloading facilities such as sidings under gantry cranes
- Further Transport Node: If the hub connects two or more transport systems, they would be joined adjacent to the transshipment system.

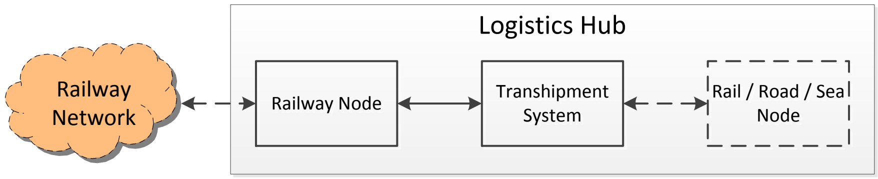

Figure 1: Logic distinction of elements in a logistics hub

Trains arriving from the railway network follow a timetable. Also transshipment systems, such as larger container terminals, use management policies to schedule train loading processes (Schönemann and Gille, 2009). In contrast, the railway node operations are carried out without a real schedule on an operational level. This is also known as the improvised operation or random schedule (Marinov and Viegas, 2011). Deviations from the timetable have consequently a negative effect on the transport times and on the productivity of railway facilities in hubs. With its improvised policy, the railway node between the rail network and the transhipment system becomes the weakest member of the transport chain where it suffers from influences from the external rail network and the transshipment terminal.

1.2 Present Rail Yard Scheduling Approach

The planning of railway capacity utilisation at intermodal terminals is carried out using simple methods. Strategic and tactical planning horizons develop schedules determining the track occupation on a daily or weekly basis. The results are track occupation plans exemplarily shown in Figure 2. These plans are prepared manually or with the help of spreadsheet software. Although, for small terminals with limited numbers of trains no special software is required, the downsides of this macroscopic planning approach become evident: Often, only one train per day is scheduled on a track. The rough plan includes large amounts of slack time. Moreover, the effective time consumption of single processes is unknown in this rough plan and rescheduling is carried out an operational basis. This leads to large amounts of idle time for infrastructure and machine utilisation.

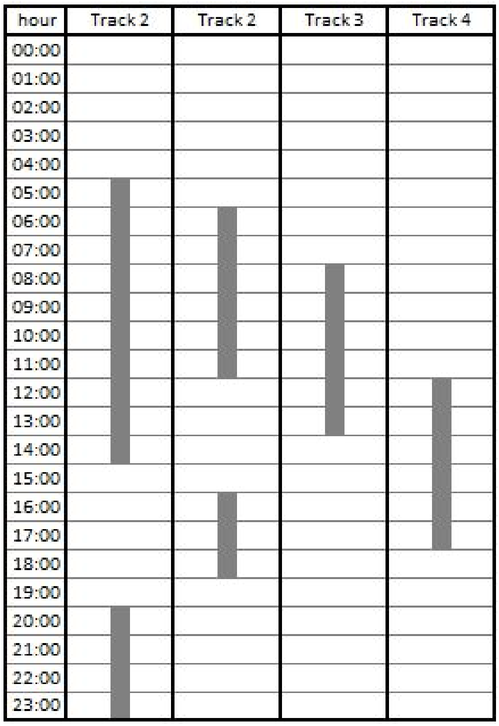

Figure 2: Exemplary track occupation plan for a small intermodal terminal with four loading tracks (24 hour scheme)

Apparently, there is a high potential of creating more efficient schedules that allow reducing the dwell time of freight wagons in the terminal. To determine this, a more detailed process analysis is required. A possible approach is presented in the following sections.

2 Methodology

In a first step, process durations have been measured in terminals of the port of Hamburg, Frankfurt (Oder) and Bologna. Whereas the port of Hamburg covers a complex and highly utilised railway infrastructure serving several rail-sea transshipment terminals, the two other locations serve as rail-road terminals that are suitable for demonstrating the methodology of the analysis.

Own measurements and data from IT systems were used to elaborate durations for railway-specific operations such as shunting or wagon break tests, and for logistics operations such as cargo inspection or transshipment. Relations between processed were identified to determine process structures and flows.

The data were used to develop distribution functions to be incorporated in a microscopic simulation. Distribution functions were developed for hub-internal processes such as dwell times of container trains in specific tracks or the transshipment of loading units. Also, a train delay function describing the deviation of incoming trains from their timetable was developed to model external influences to the hub schedule.

The result of this methodology is a combined railways and transshipment hub model implemented using the event-based simulation environment Simul8. For the most detailed results it was favourable to implement a microscopic modelling approach which allows separating the node into its single components. The components of the node are thought of as interconnected queuing systems that interact and influence each other, so that the global impact of freight train operations in a rail network context is captured.

3 Modelling Approach

The capacity of railway infrastructures can be measured analytically using the UIC Code 406 compression method (International Union of Railways (UIC), 2004). The method should allow the user "to carry out capacity calculations - following common definitions, criteria and methodologies from an international standpoint - for lines/nodes or corridors" (International Union of Railways (UIC), 2004) by using a standardised compression method. During the compression, the blocking time stairways of timetable have to be shifted together as close as possible. This allows calculating the occupancy rate as the difference between the evaluated time period and the time which is elapsed by the blocking times of the existing train paths. Finally, the occupancy rate has to be compared with limit values predicted in UIC Code 406 (c.f. Lindner, 2011).

Even though the UIC Code 406 method enables the planner to determine line capacities quite well, it cannot be applied for node capacity research. Lindner applies the method on railway stations and explains "why the occupancy rate cannot provide a significant parameter for node station capacity". To determine a railway node's productivity it is rather required to determine the capacity utilisation of its assets, e.g. the length of occupancy of tracks or the transshipment time per train under the crane.

Microscopic railway simulation tools provide good help when analysing railway infrastructures. Logistics processes, however, cannot be mapped by them. But, the tools provide valuable data that can be used as input parameters in further research. In this paper, the mapping for railway-related and logistics processes is carried out using the event-based simulation environment Simul8 (Concannon, 2003) which permits recognising relationships between processes, their durations and infrastructure occupation times. Some of its simulation parameters could be gained from microscopic railway simulations beforehand.

The model of a road-to-rail intermodal terminal has been set up in the simulation environment as illustrated in Figure 3. It consists of a small marshalling yard of four formation tracks and two loading tracks for cargo transshipment. The model performs railway operations such as shunting or train making and loading/unloading processes. The formation tracks perform processes such as train inspection, brake test etc. One shunting

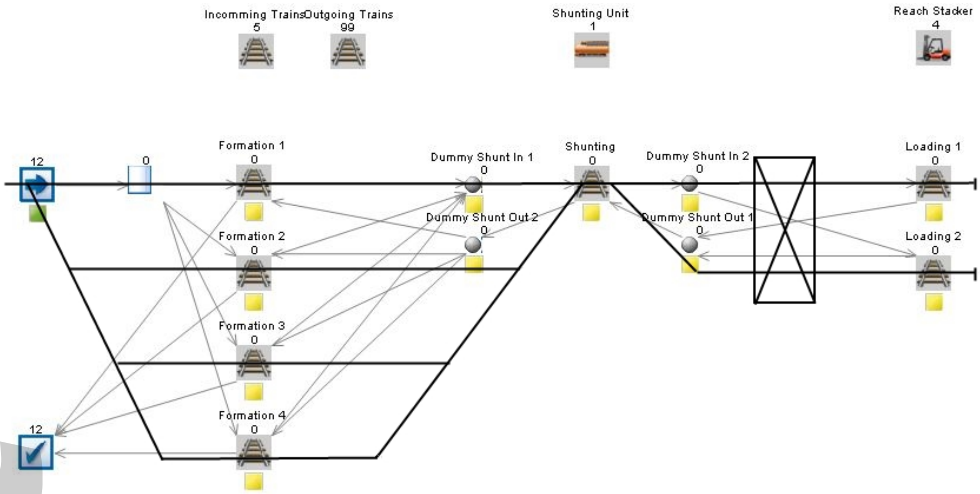

Figure 3: A simple intermodal terminal in Simul8 with two loading tracks and a small marshalling yard of four formation tracks

locomotive operates at the modelled yard. Its operational properties and driving dynamics were taken from a prior microscopic railway simulation of the marshalling yard. The transshipment processes at the loading tracks are operated by reach stackers.

As illustrated in Figure 3, the railway infrastructure in a terminal is usually divided into two parts:

Formation Tracks: These tracks hold incoming and outgoing trains. They are used for specific processes such as coupling of wagons, brake test, or inspection of freight.

Loading Tracks: The tracks are located in the terminal itself and used for loading and unloading freight wagons. Specific equipment such as reach stackers or container crane are used for the transshipment processes.

4 Analysis and Computational Results

4.1 Process Sequencing in Rail Yards

The various processes a train has to pass through are interconnected and subject to the process control. The simulation environment allows to regard

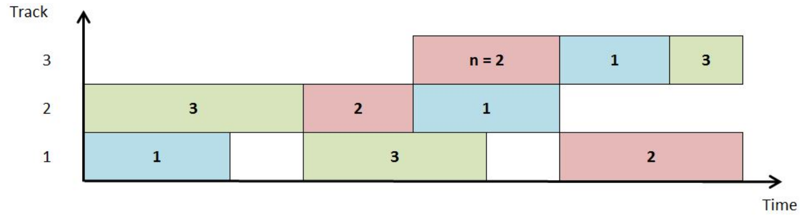

Figure 4: Graphical illustration of a Job-Shop Scheduling Problem

the process chain as a queuing system. Each track acts as a queuing station. A station in terms of the simulation can have several states:

- Waiting (track is empty, idle),
- Working (a train is being processed, e.g. loaded/unloaded),
- Blocked (train processing is completed but train cannot leave due to an infrastructure conflict),
- Resource starved (resources unavailable, e.g. crane not ready).

For specific processes, staff and machinery is employed. The inspection of incoming cars for example, requires the employment control teams. Loading and unloading requires the provision of handling equipment and the appropriate operators.

Trains pass through the model adressing different stations in a specific sequence. The process chains can be studied with methods of scheduling in order to develop robust track occupation plans. The task of the scheduling is the assignment of n jobs (trains and freight cars) to m machines (tracks). The processes must be executed in accordance with a specific order restriction. It is also necessary to make this mapping optimally with respect to a given objective (e.g. the minimum total cycle time). This has been be achieved solving a Job-Shop-Scheduling-Problem (JSSP).

The Job-Shop Scheduling Problem (JSSP) is one of the classic scheduling problems in operations research. It's role is to assign jobs to machines and optimise this allocation under various boundary conditions, where the aim is to carry out all orders as quickly as possible. The JSSP creates job plans for the formation and the loading tracks. Figure 4 illustrates a job plan for three freight wagons passing through a terminal with three stations.

4.2 Theoretic Capacity and Marginal Efficiency

Simul8 uses dynamic discrete simulation building a queuing system to model flows of processes. The model allows any conclusions on the throughput behaviour of the stations, possible bottlenecks, and handling strategies.

Each element of the model can be understood as a station $i \in I$ of a queuing system. The service rates μ_i for each station i have been determined by empirical studies and are expressed as distribution functions for each station. Another crucial variable is the arrival rate λ which represents the intermediate arrival time where trains are entering the system.

In the initial simulation scenario λ and μ_i are kept constant in order to determine the theoretic throughput capacity of the overall hub system. It can be expressed as the service time of the whole system s. In further scenarios stochastic impacts are introduced to determine the impact of disturbances on the hub's productivity.

In order to determine the hub's overall maximum capacity and the effects of overload, a first simulation is run using deterministic (fixed) service rates on all stations. The arrival rate $\lambda = \frac{1}{E_i}$ with $E_i = const$ (stochastically independent and equally distributed) was stepwise decreased during several simulation runs in the interval $180 \geq E_i \geq 60$ min. As a measure of quality the average time of trains in the system $\bar{t} = \frac{1}{n}\sum T_n$, with $n \in N$ as the set of all trains and T_n as the overall time of train n in the system, was monitored.

Figure 5 illustrates the results of the simulation: When the intermediate arrival time λ is low, $\bar{t}$ is stable on a low level. When rising λ, more trains arrive at the hub per time period. Thus, $\bar{t}$ increases at a certain point due to hub-internal congestion effects, which can still be caught and do not cause congestion phenomena outwards. If one increases λ further so that $\lambda > \bar{t}$, trains pile up in front of the hub (Average Time in Queue $>$ 0 in Figure 5). Now $\bar{t}$ rises to a higher but constant level. Consequently, overload does not only lead to congestion phenomena on railway network structures outside the terminal. Also internal blockades occur that result in slower processing of trains through the hub. The throughput time of trains or wagons (and cargo) through the hub increases.

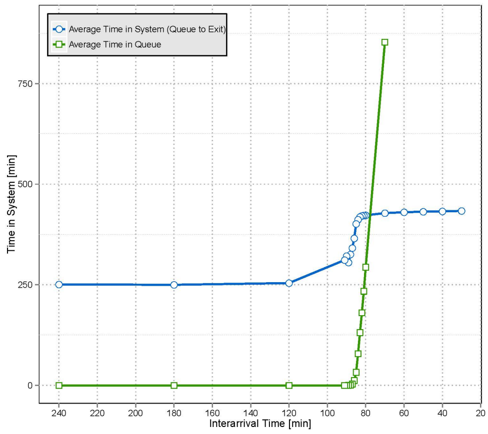

Figure 5: Determination of the terminal's performance limit: Congestion appears at the terminal entrance when the arrival rate is high (green line). Overload results in an overall slower processing and has a negative effect on the train handling time.

4.3 Detection of Bottlenecks

While the previous section investigated the hub as a whole, it is useful to examine the performance of single stations in detail to discover bottlenecks. Figure 6 illustrates the utilisation of a loading track in the model subject to the arrival rate λ. The utilisation of the track increases with λ (green line). Simultaneously, the idle time decreases (blue line). In the left graph the utilisation rate remains at about 45% but resource starvations arise (yellow line). In these cases it was identified that no reach stacker was available while trains were waiting to be served. By employing an additional reach stacker, it was possible to raise the utilisation rate of the track and reducing the resource starvations (right graph of Figure 6). However, now blockades (red line) occur which indicate a bottleneck on another infrastructure element of the hub. It can be solved by analysing other hub elements.

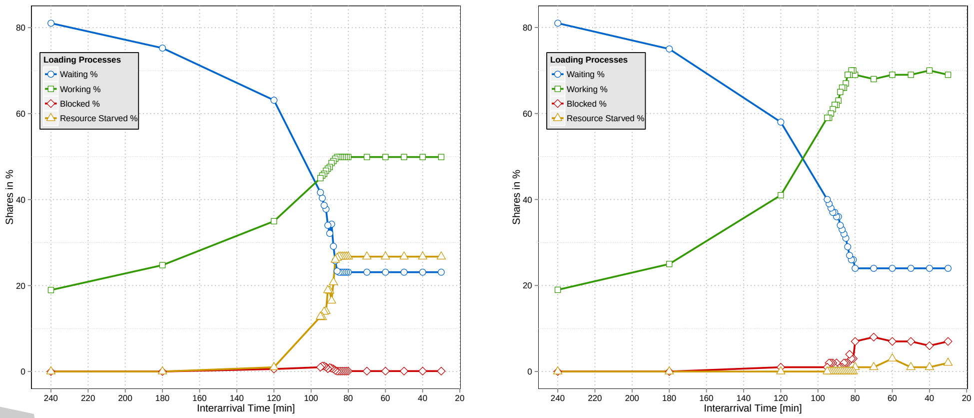

Figure 6: Utilisation analysis of a loading track. Left: Resource starvations occur because transhipment facilities are not ready. Productivity is low. Right: Increasing transhipment capacity allows higher productivity but causes another bottleneck elsewhere.

5 Stochastic Effects and their Impact on the Network

5.1 Process Duration Uncertainty

In section 4 the service rate μ_i and the arrival rate λ were fixed to an average value for each station i. An initial job schedule with twelve trains calling the hub was created. Train arrivals were distributed equally over time and processing durations at the stations were constant. In this ideal process flow scenario (see Figure 7) idle time could be minimised and occupancy conflicts be avoided. The result illustrates the best (theoretic) job schedule for the simulation model.

The best possible utilisation of handling equipment is rarely feasible and not very realistic. In order to examine stochastic effects and their impacts in the hub's productivity, distribution functions are used in the further examination. Distribution functions were computed from observations at intermodal terminals for all stations of the simulation accordingly. This enables to analyse hub-internal process deviations and their impacts on the transport chain. Figure 8 shows exemplarily the distribution function used for the operations of loading tracks.

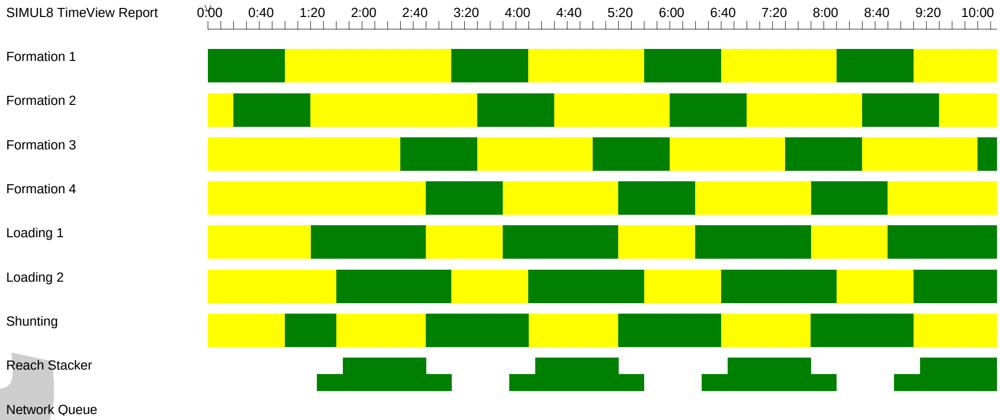

Figure 7: Occupation plan for all tracks in the first simulation scenario, $\lambda \approx 0.72$ train arrivals per hour.

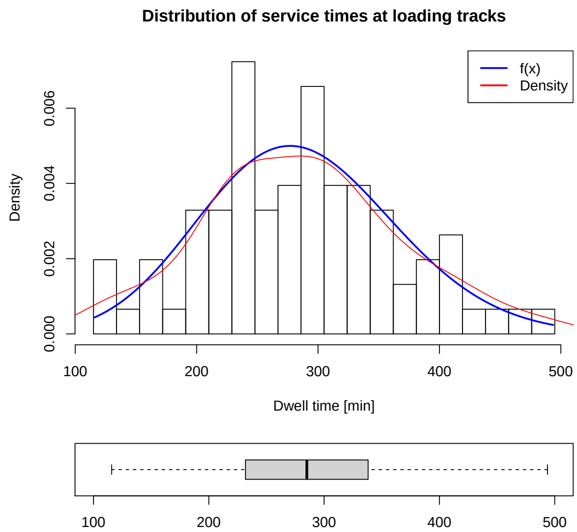

Figure 8: Distribution function of train service times in loading tracks

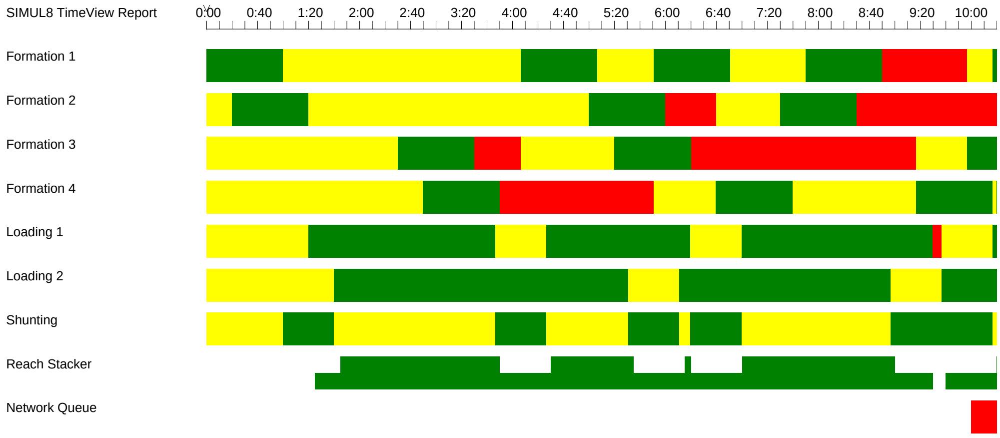

Figure 9: Occupation plan for all tracks in the last simulation scenario, $\lambda \approx 0.25$ train arrivals per hour.

Stochastic data were inserted stepwise to create different scenarios. The last scenario, considering the most stochastic effects, is shown in Figure 9. Apparently, numerous track allocation conflicts (red blocks) occur. Thus, the processing time of the 12 planned trains raised from about 17 hours (initial scenario of Figure 7) to about 22 hours. Delayed trains generate further delays in the rail yard but have also an impact on the machinery and crew planning.

In order to avoid such allocation conflicts, process delays need to be incorporated in the schedules. The usual way, the integration of buffer times, but again leads to an extension of the train dwell time. It is therefore of particular interest to predict the reliability for the proposed schedules in advance to minimise the time buffers required.

Carey (1999) proposes an ex ante method using probabilities of delays as a measure of reliability. The method has been developed for schedules in public transport but can be transferred to freight transport schedules. Therein, the probability that a train t arrives or departs at a regarded station after its scheduled time m_t is given by $\int_{-\infty}^{m_t} d\tau$.

For more in-depth information on this topic, please refer to the original paper. At this point it shall be only noted that the probabilities calculated by this method can be incorporated in the simulation model and thus enable the creation of reliable schedules.

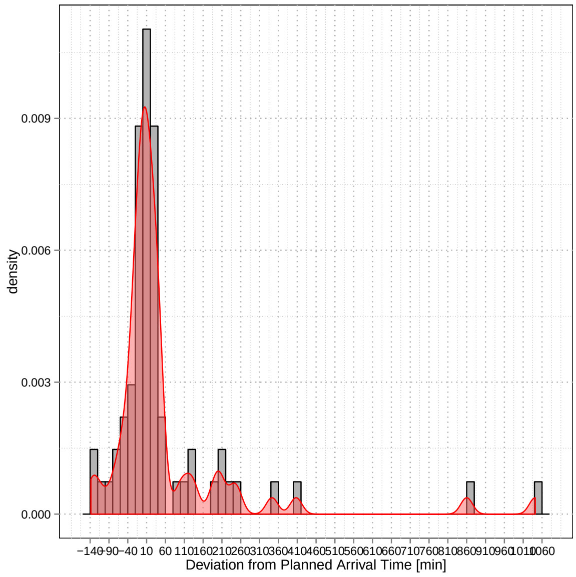

Figure 10: Sample of freight train arrival deviations on a network cross-section.

5.2 Train Arrival Uncertainty

Train arrival uncertainty is one of the most important constraints to planning effectiveness in container terminals, especially in operational (short-term) planning. In mixed railway networks, where freight trains have a lower priority than passenger trains, deviation from the timetable are prevalent. Late as well as early arrivals up to several hours, as shown in the sample of Figure 10, are possible.

Likewise to the process duration uncertainty, the delays of arriving trains were integrated as stochastic effects to the simulation model. This was realised using the achievements of (Wendler and Naehrig, 2004), where the authors describe a method for the statistical analysis of train delay data.

The reliable estimation of train arrival times is the key for successful scheduling in intermodal terminals. An increase in terminal productivity can be achieved by using methods of rescheduling at the operational level when the estimated time of arrival (ETA) for a train is known.

The current provision of ETA is not reliable due to unexpected circumstances in freight train forwarding. However, there are some approaches developing to predict the ETA and to use these information in intermodal terminals and other freight nodes. Based on the findings so far, the concept is currently being further developed by the author.

6 Summary

Intermodal transshipment hubs cannot perform their maximum capacity utilisation due to effects of internal and external disturbances on the process flows. The presented approach allows identifying influences of process flow disturbances on individual elements of the hub and the detection of bottlenecks within the hub. Congested situations result in slower throughput of trains through the node but also reduce effective available infrastructure capacity.

Identifying the infrastructure allocation conflicts is carried out by microscopic event-based simulation and the definition of appropriate train schedules replacing the currently used improvised random schedule. These new schedules are de-congested by introducing reliable buffer times based of delay probabilities. Another strong impact on the schedule feasibility have delayed train arrivals. They can be tackled by an efficient rescheduling algorithm combined with methods for the reliable calculation of the estimated time of train arrival (ETA) for incoming trains.

References

Malachy Carey. Ex ante heuristic measures of schedule reliability. *Transportation Research Part B: Methodological*, 33(7):473–494, 1999. ISSN 01912615. doi: 10.1016/S0191-2615(99)00002-8.

Armando Carteni and Stefano De Luca. Simulation of a container terminal through a discrete event approach: literature review and guidelines for application. In Association for European Transport, editor, *European Transport Conference*, 2009.

Kieran Concannon. *Simulation modeling with SIMUL8*. Visual8 Corp., Mississauga and Ont., 2003. ISBN 0-9734285-0-3.

International Union of Railways (UIC). *UIC-Code 406: Capacity*. Paris, 2004. ISBN 2-7461-0802-X.

Tobias Lindner. Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity. *Journal of Rail Transport Planning & Management*, 1(1):49–57, 2011. ISSN 22109706. doi: 10.1016/j.jrtpm.2011.09.002.

Marin V. Marinov and José M. Viegas. A mesoscopic simulation modelling methodology for analyzing and evaluating freight train operations in a rail network. *Simulation Modelling Practice and Theory*, 19(1):516–539, 2011. ISSN 1569190X. doi: 10.1016/j.simpat.2010.08.009.

Andrew Nash and Daniel Hürlimann. Railroad simulation using OpenTrack. In John J. Allan, editor, *Computers in railways IX*. WIT Press, Southampton and England and Boston, 2004. ISBN 1-85312-715-9.

René Schönemann and Andreas Gille. Kapazitätsreserven und Betriebsabwicklung bei der Hafenbahn Hamburg nach Umstellung auf die Fahrplanfenster-Strategie Slotmanagement (unpublished), 2009.

Ekkehard Wendler and Michael Naehrig. Statistische Auswertung von Verspätungsdaten. In Verband Deutscher Eisenbahn-Ingenieure e.V., editor, *Eisenbahningenieurkalender*, pages 321–331. Tetzlaff - Verlag, Hamburg, 2004.