

Modeling landside container terminal queues: Exact analysis and approximations

Debjit Roy ^{a,b,*}, Jan-Kees van Ommeren ^c, René de Koster ^b, Amir Gharehgozli ^d

^a Indian Institute of Management Ahmedabad, India

^b Rotterdam School of Management, Erasmus University, The Netherlands

^c Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, The Netherlands

^d David Nazarian College of Business and Economics, California State University Northridge, USA

ARTICLE INFO

Keywords:

Bulk arrivals
Synchronization queues
Priority queues
Shared resources
Container terminal
Landside design
Queuing models

ABSTRACT

With the growth of ocean transport and with increasing vessel sizes, managing congestion at the landside of container terminals has become a major challenge. The landside of a sea terminal handles containers that arrive or depart via train or truck. Large sea terminals have to handle thousands of trucks and dozens of trains per day. As trains run on fixed schedule, their containers are prioritized in stacking and internal transport handling. This has consequences for the service of external trucks, which might be subject to delays. We analyze the impact of prioritization on such delays using a stochastic stylized semi-open queueing network model with bulk arrivals (of containers on trains), shared stack crane resources, and multi-class containers. We use the theory of regenerative processes and Markov chain analysis to analyze the network. The proposed network solution algorithm works for large-scale systems and yields sufficiently accurate estimates for performance measurement. The model can capture priority service for containers at the shared stack cranes, while preserving strict handling priorities. The model is used to explore the choice of different internal transport vehicles (with coupled versus decoupled operations at the stack and train gantry cranes) to understand the effect on delays. Our results show that decoupled transport vehicles in comparison to coupled vehicles can mitigate the external truck container handling delays at shared stack cranes by a large extent (up to 12%). However, decoupled vehicles marginally increase the train container handling delays at shared stack cranes (up to 6%). When train arrival rates are low, prioritizing the handling of train containers at the stack cranes significantly reduces their delays. Further, such prioritization hardly delays external truck containers.

1. Introduction

Global container throughput is growing strongly. It is projected to increase from 650M TEU (twenty-foot equivalent unit) in 2013 to 973M TEU in 2023.¹ Container terminals play an important role in the global trade and act as hubs for intermodal transport. The seaside of a terminal handles containers that arrive (import) or depart (export) via vessels whereas the landside handles containers that arrive (export) or depart (import) via trucks, trains, or barges.

* Corresponding author at: Indian Institute of Management Ahmedabad, India.

E-mail addresses: debjit@iima.ac.in (D. Roy), j.c.w.vanommeren@utwente.nl (J. van Ommeren), RKoster@rsm.nl (R. de Koster), amir.gharehgozli@csun.edu (A. Gharehgozli).

¹ <https://www.seatrade-maritime.com/asia/global-container-port-volumes-forecast-973m-teu-2023-drewry>.

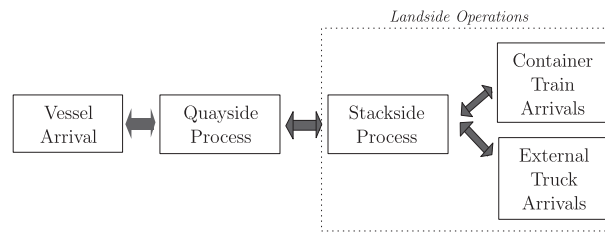


Fig. 1. Landside operations with export and import flows.

Recent trends in ocean transportation pose several operational challenges at both the seaside and the landside of a terminal. The introduction of large container ships such as the MSC Gulsun, which can carry over 23,000 TEU has put significant pressure on ports to develop the infrastructure to handle such ships. At the seaside, many projects are underway to increase the vessel draught and throughput capacity. However, at the landside of the terminal, thousands of trucks and multiple trains have to be handled in a very short time span. It is difficult to cope with the sudden and massive peak in throughput requirements caused by such large vessels. For example, the Loadstar reports that “the UK’s biggest container port of Felixstowe has been challenged by a surge of ultra-large container vessels (ULCVs) that require more gangs and more cranes to service the increased cargo exchange of regularly more than 5000 boxes per call. This in turn exerts pressure on the landside operation in a vicious circle of reduced port productivity”.²

The high variability in container arrivals puts another challenge on the landside operations. ULCVs frequently arrive outside their official schedule windows. This results in a bunching of big ships all vying for the same berths at the same time. In turn, greater variability in ship arrivals adds flow variability at the landside. Any delay at the landside operations has a cascading effect on the timeliness of the downstream hinterland connections and may affect terminal performance as a whole. Due to landside congestion, at times twelve container ships can be anchored in the waters off the ports of Los Angeles and Long Beach.³ Hence, denser container traffic and greater variability in daily volumes are increasingly causing longer delays at the terminal landside. In addition, the landside area, which faces city dwellers, is often constrained by geographical expansion limits.

Landside processes are often the source of long operational delays at the terminal. Container dispatch delays can occur both at the terminal gates and inside the terminal. In response, terminals have adopted terminal appointment systems (TAS) for trucks and they have introduced incentive schemes to level the truck traffic across the day. Studies reveal however that, although a TAS can reduce truck congestion at the terminal gates, it cannot prevent internal delays. For example, harbor truckers at Los Angeles Beach, which uses a TAS, still continue to experience long delays at the terminals. The data reveal that the worst delays are not spent waiting at the terminal gates, but rather inside the terminals (average delay of 19 min at the gate vs. 71 min inside the terminals). The delays were attributed to chassis shortages.⁴

We develop a stochastic model to address the congestion problem caused by *export containers* that arrive at the terminal via two modes of transport: external trucks (ETs) and container trains. Fig. 1 illustrates the scope of this research.

Containers arriving via ETs wait for service inside the terminal in buffer positions at the stack blocks, typically operated by automated stack cranes (ASC). The trains run on fixed schedules and have to depart on time. They therefore take service priority over the ETs. The containers brought by trains are transported to the storage stacks by (internal) terminal vehicles. There are two types of terminal vehicles (see Fig. 2): (1) coupled, i.e., human-operated terminal trucks with trailers, or automated guided vehicles requiring hard-coupling (synchronization) with both the stack and train gantry cranes to load or unload the containers on or off the vehicle bed and (2) decoupled, i.e., lifting vehicles (LVs), both human-operated reach stackers and automated lifting vehicles (ALVs) that can dropoff (pick up) containers on (from) a container frame or the ground and can therefore operate decoupled from the cranes. Currently, mostly coupled vehicles are used, as seen for example at the Port of Long Beach in LA (USA), at the Maasvlakte 1 terminals in the Port of Rotterdam (the Netherlands), and at JNPT Port (India). We refer to a coupled terminal vehicle for train container movement as a terminal truck (TT). Although decoupled vehicles have a higher throughput capacity compared to coupled vehicles, they are also more expensive. In this research, we compare container throughput time performance of coupled systems and decoupled systems. In order to make sure trains meet their schedules and decongest the space around the GC, container terminals prioritize train containers over ET containers at the shared resource — the ASCs. Although this priority reduces train container throughput times, it can also lead to excessively long throughput times of ET containers. This not only affects the delivery reliability to the beneficiary cargo owner, but also reduces the number of trips for truck drivers, leading to shortage of driver capacity.⁵

The operational steps for the train unloading process using TTs include: (1) The TT first queues at the trainside to pickup the container present on the train, (2) The GC unloads the train container and drops it off on the TT, (3) The TT then travels with the container and queues at the stack, (4) The ASC picks the container from the TT and the TT dwells at the nearby buffer lane waiting to process the next transaction. Hence, the TTs are coupled with the GC on the trainside and the ASCs at the stackside.

² <https://theloadstar.co.uk/congestion-felixstowe-pushes-maersk-lavras-london-gateway-boosting-asia-europe-call-hopes/>.

³ https://www.joc.com/port-news/la-long-beach-container-ship-backup-reaches-2-year-high_20141111.html.

⁴ https://www.joc.com/port-news/terminal-operators/delays-la-lb-truckers-worst-inside-terminal-not-gates_20150909.html.

⁵ <https://www.wsj.com/articles/truck-driver-shortage-supply-chain-issues-logistics-11635950481>.

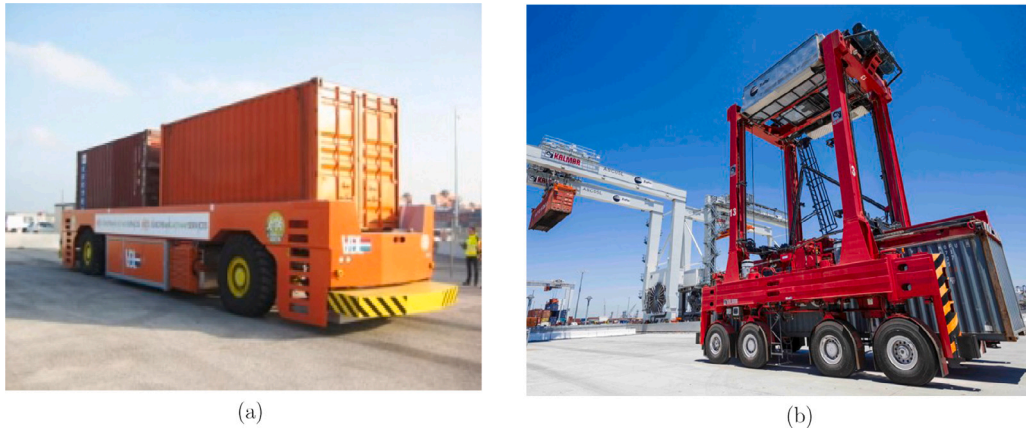


Fig. 2. Terminal trucks for internal transport between container train and stack block: (a) a coupled automated guided vehicle (AGV) cannot lift up (set down) a container directly from (on) the ground (source: <http://www.weweler.eu/nl/vdl-scoort-miljoenenorder-voor-nieuwe-agv/>), (b) a decoupled automated lifting vehicle (ALV) can lift up (set down) a container directly from (on) the ground (source: Kalmar AutoStrad™ <https://www.kalmarglobal.com/equipment/straddle-carriers/autostrad/>).

The operational steps for the train unloading process using LVs include: (1) The GC unloads the train container at a train buffer lane. The container now queues at the train buffer lane for internal transport to the destination stack. (2) An LV arrives at the train buffer lane, picks up the container and transports it to the destination stack buffer lane, (3) The container now queues for the ASC availability, and the ASC then picks up the container and stows it at an appropriate storage location. It is not clear how the coordination delays introduced in the coupled system increases the required number of TTs for a desired throughput level. It is also not clear how the coupled nature of TTs affects the ET waiting times at the ASC.

Although some research has focused on the design of container terminal seaside operations, studies that analyze landside operations are limited. Those that do, do not focus on the congestion of vehicles at the shared resources (see Carlo et al. (2013), Gharehgozli et al. (2015), Zhen (2016), Roy and De Koster (2018), Dhingra et al. (2018), Roy et al. (2020)). Dhingra et al. (2018) developed a container terminal model including truck operations with time-varying truck arrival rates. Some other papers that also examine landside operations do not really focus on the internal operations of container terminals (see e.g., Giuliano and O'Brien (2007)). The interactions between ET and TT containers at the ASCs within the terminal have not been explicitly modeled before. We explicitly model this interaction between the ETs and TTs at the ASCs. The research questions we address include:

- Trains arrive at the landside with a large number of export containers (single-stack or double-stack). Hence, the arrival process of train containers can be modeled as a bulk arrival process. How can we formulate a queuing network model for landside operations at a container terminal for a (de)coupled system with bulk arrivals and non-preemptive handling priority for the TTs over ETs at the shared stack crane? How do we evaluate the network and obtain performance measures?
- What is the effect of load prioritization and vehicle synchronization (coupled or decoupled) on the external truck vs. train container handling delays at the ASCs?

To analyze the processes at the terminal landside, we develop a multi-class semi-open queuing network (SOQN) model with automated stack cranes (ASCs) and train container gantry crane (GC) as key stations, and independent container arrival streams at different stations. The model captures the congestion during container handling at the train GC and ASCs, and delays associated with vehicle movement between the GC and ASCs. Our network can be classified as a semi-open queue with multiple-customer classes, two arrival streams at the shared resources (stack cranes) with non-preemptive priorities of customer classes, and general service times. Unfortunately, analytic solutions for such networks are not available. Existing SOQN models only allow for a single stream of customer arrivals (e.g., see Jia and Heragu (2009)). In addition, existing SOQN models do not account for bulk arrivals and shared resource queues, which is a modeling opportunity.

While developing simulation model is another possibility, running multiple simulation scenarios typically requires developing different simulation models (e.g. for coupled and decoupled vehicles), which can be very time-consuming. Analytical modeling offers advantages over simulation. First, it gives more structural insights in relations between parameters and in stability conditions for the network. Second, it helps in the terminal design conceptualization phase because they allow for inexpensive enumeration of terminal design parameters to find appropriate designs.

We start with coupled resources and propose a three-step modeling and solution approach (see Fig. 3). In Step 1, we analyze a single shared resource (ASC) in isolation with two customer streams: (1) external truck (ET) arrivals at the shared resource, and (2) state-dependent arrivals from a finite source (the TTs), and non-preemptive priorities at the shared resource. In Step 2, we analyze multiple of these shared resources operating in parallel i.e., multiple ASCs handling containers from both ETs and trains in parallel. In Step 3, we include the synchronization station with two buffers in the network, to match waiting containers (arrived in

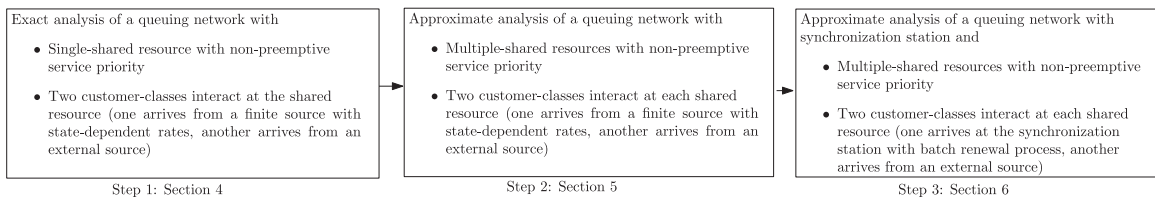


Fig. 3. A three-step stochastic modeling and analysis approach for a coupled system.

a train) with a TT. The train containers with bulk arrivals queue at the first buffer whereas the idle vehicles waiting to transport train containers queue at the second buffer.

We are able to exactly analyze the network for a special case, i.e., when container trains are always available (infinite source) and the network includes a single shared resource. However, for the general case when the TTs may have to wait for container trains to arrive and in the case of more than one shared resource, the network is intractable. For this case, we develop an approximate approach to evaluate the network performance measures. We also derive network stability conditions for both simple and complex network configurations. Using discrete-event simulation, we show that the approximate analysis captures the dynamic interactions between container train arrivals and truck arrivals at the ASCs quite well, particularly for large-scale problem instances.

The paper makes the following contributions: (1) *Domain contribution*: we model the interactions between containers arriving by high-priority trains, external trucks and shared stack cranes and show the impact of prioritization and vehicle type on the throughput time performance for train and truck containers. This problem is relevant and has not yet been given attention in research. Managers may find the model helpful to decide on coupled (TTs) and decoupled (LVs) transport vehicles. (2) *Theoretical contributions*: we provide an exact analysis for a closed queuing network with two customer classes with service priority at the shared resource. The customers belonging to the higher priority class arrive from a finite source, whereas the other class customers arrive from an external source. For such networks, we derive conditions for network stability. For the larger, more complex network with multiple shared resources, we develop approximate solution methods that perform particularly well for medium to large problem instances. We model and analyze the system as a semi-open queuing network with batch arrivals, external arrivals at the shared resources, and service priorities at the shared resources for the large network.

The rest of the paper is organized as follows. In Section 2, we review literature on stochastic models in container terminals and identify the gap in the literature. We first focus on coupled systems and describe a typical landside terminal system, state our modeling assumptions, and present the integrated train-truck queuing network model in Section 3. We evaluate the coupled system performance with one ASC, multiple circulating TTs, and ET arrivals in Section 4. We analyze the coupled system performance with multiple ASCs, multiple circulating TTs, and ET arrivals in Section 5. Finally, we estimate the performance measures for coupled system with train arrivals at the synchronization queue in Section 6. We validate our model and present the model insights with different scenarios in Section 7. We state our conclusions and discuss scope for further research in Section 8.

2. Literature review

We review the literature on landside container terminal operations and also motivate the usage of stochastic models for performance analysis of landside operations from related studies in seaside container terminal operations. A brief review on the literature shows that in general, most papers focus on a specific subsystem of a container terminal in isolation (see, for example, Gorman et al., 2014; Gharehgozli et al., 2015, 2020). However, there is a need for integrated and transparent solutions that can contribute to enhancing service efficiency (Li et al., 2021, 2020b; Wang et al., 2021a; Duru et al., 2020; Yang, 2019). Landside decisions made in isolation can be categorized into external truck (ET) operations and train operations. We first discuss studies on external truck as well as train operations. We then focus on integrated decision areas.

2.1. Truck operations

Several researchers study truck flow control strategies at the landside that can be implemented using the Terminal Appointment System (Torkjazi et al., 2018; Wibowo and Fransoo, 2021). There are several studies on determining the maximum number of trucks allowed at a given hour. For example, Mar-Ortiz et al. (2020) design an optimization-based decision support system (DSS) to balance supply (equipment assignment) and demand (arrivals of trucks) while improving customer service level and productivity indicators. They determine the appointment quota for each time slot on a one-day planning horizon. Using a mixed integer linear programming model, Zehendner and Feillet (2014) address the joint decision problem of determining the number of truck appointments to offer per time slot and allocating the straddle carriers to different transport modes at the landside. Azab and Morita (2022) propose two binary integer programming models to study how to stack container according to the schedule set by the truck appointment system to minimize the number of relocations. Chen et al. (2021) study an autonomous truck based scheduling problem for container transshipment between two seaport terminals (also see You et al., 2020). They consider that autonomous trucks travel in platoons with a short inter-vehicle distance. They formulate the problem as a mixed-integer second-order-cone programming model to minimize the total operation cost. They solve the problem using a column-generation-based heuristic method. Li et al. (2020a) study

the truck appointment system at multi-terminal ports where each terminal normally operates its appointment system independently. Therefore, there is a lack of coordination on how time slots are allocated to trucks across multiple systems. To address the issue, they work on a game theoretical framework that studies the exchange of Tradable Truck Permits across terminals.

Queuing models (both stationary and non-stationary) are also used to manage congestion at the terminal gates. Guan and Liu (2009) develop a multi-server queuing model to analyze gate congestion and quantify waiting costs. Chen and Yang (2010) determine the time-windows that minimize transport costs, including waiting costs, fuel consumption, storage time, and yard fee. Such time-windows flatten the peaks of truck arrivals. Giuliano and O'Brien (2007) evaluate the effect of a gate appointment system and off-peak operating hours on reducing queues at the gates. Phan and Kim (2016) propose a negotiation framework based on mathematical programming model to optimize truck arrival times at peak times. The model also attempts to minimize both truck waiting cost as well as the truck inconvenience cost so that the deviation from their preferred times is minimal.

Note that most studies do not consider the interaction of external trucks with other modes of transport such as trains at the stack blocks. We contribute to the literature by working on an integrated problem.

2.2. Train operations

Research on train operations can be categorized to several streams that are mostly focused on transshipment yards where containers are moved between trains and trucks (Schulz et al., 2021; Rupp et al., 2021). For an overview of relevant research questions, one can refer to Berndörfer et al. (2018) or Boysen et al. (2012a). All in all, the most fundamental problem is the train scheduling problem which determines the route, stop plan and departure/arrival times of each train (see, for example, Zhang et al., 2022; Ahern et al., 2022; Wang et al., 2021b; Frisch et al., 2021). In a different segment of the literature, some papers study the added value of a rail terminal located inside a port area (Xie and Song, 2018; Hu et al., 2019; Chen et al., 2018; Gharehgozli et al., 2017).

In another avenue for research that is more relevant to this paper, a portion of the literature study rapid rail–rail transshipment and develop optimization models to unload and reload containers (Jaehn, 2013; Boysen and Flidner, 2010; Boysen et al., 2011, 2012b). In such problems, maximizing the rolling stock efficiency, optimizing the aerodynamics of intermodal freight trains, and optimally assigning containers to wagons (i.e., train load planning for single stack or double stack trains) are among the main objectives (Ng and Talley, 2020; Upadhyay, 2020; Mantovani et al., 2018; Upadhyay et al., 2017; Ambrosino and Siri, 2015; Lai et al., 2008). In this regard, an interesting area of research is the concept of blocking which assigns railcars to blocks remaining in a group over a long distance moving from one yard to another (Ruf and Cordeau, 2021). Bruck et al. (2021) study the blocking problem (also known as classification) and propose mixed integer programming models to decide how inbound trains are split into sequences of railcars, on which tracks these railcars are handled, and how the railcars are assigned to outbound blocks.

Queuing models have been also used to study train operations. Again, most papers focus on network design, traffic flow, and capacity planning problems (Bychkov et al., 2021; Huisman et al., 2002; Weik et al., 2016; Huisman and Boucherie, 2001; Leachman and Julia, 2012). The train classification process has been analyzed using different queuing models as well (Dorda and Teichmann, 2013; Turnquist and Daskin, 1982).

We contribute to the scarce literature of train handling operations inside a container terminal by modeling the unloading operations using different types of vehicles.

2.3. Integrated operations

Research on integrated terminal decisions are mostly focused on the seaside operations. Methodologies include both optimization and queuing models.

Most optimization based seaside planning decisions such as berth allocation, quay crane assignment, and storage space allocation are planned in a sequential manner. However, a sequential decision making can result in low resource utilization and high operational costs. Meisel and Bierwirth (2013) propose a three-stage framework for integrating all decisions. For example, the first phase estimates crane productivity rates which is used in the second phase to determine berthing decision and assigning crane capacity to the vessels. Scheduling decisions are made in the third phase. Liu et al. (2016) develop a bi-level optimization model to jointly allocate the berths to the vessel and allocate yard space to the vessel unloading operations. While the objective for the berth allocation is to minimize the deviation from the vessel expected turnaround time window, the objective for the yard space allocation is to minimize the container internal transportation time. Kizilay et al. (2020) develop multiple constraint programming models to optimize the integrated equipment assignment and scheduling problem at the seaside. The objective is to minimize the turnover times of the vessels and maximize terminal throughput. (Vacca et al., 2013) develop an exact branch and price algorithm to simultaneously optimize allocating berth and assigning quay crane in seaport container terminals. Wang et al. (2018) develop a column generation procedure to solve the large-scale integrated optimization problem comprising of berth allocation, quay crane assignment, and yard storage space to the incoming vessels. Liu et al. (2021) study the integrated planning problem of berth allocation and vessel sequencing in a port with a one-way navigation channel (also see Corry and Bierwirth, 2019).

Stochastic models have analyzed congestion issues at the seaside operations. Hoshino et al. (2007) is among the pioneers to develop a queuing model for terminal operations. Kang et al. (2008) develop a cyclic queue model to estimate the steady-state port throughput and optimize fleet size at the terminal using generally distributed crane service times and truck travel times. Roy et al. (2020) develop integrated queuing network models to analyze the container throughput time performance for a terminal with the quay crane (QC) operating in a single mode. They also generate insights with respect to the vehicle dwell point strategies

Table 1
Comparing this paper with the literature.

Scope	Author	Container source	Container arrivals	Transport vehicle	Interactions at ASC	Model	Solution method
Seaside	Hoshino et al. (2007)	Vessel	Single	Coupled	No	Closed queueing network	Convolution
Seaside	Lee et al. (2014)	Vessel	Single	Coupled	No	Analytical	Markov chain analysis
Seaside	Roy and De Koster (2018)	Vessel	Single	Decoupled	No	Semi-open queueing network	Decomposition
Seaside	Roy et al. (2020)	Vessel	Single	Coupled, Decoupled	No	Semi-open queueing network	Decomposition
Inter-terminal	Mishra et al. (2017)	Vessel, Barge, Train, Truck	Single	Coupled (heterogeneous capacity)	No	Semi-open queueing network	Network decomposition
Landside	Dhingra et al. (2018)	Truck	Single	Coupled	No	Semi-open queueing network	Matrix Geometric Method
Landside	This paper	Train, Truck	Bulk	Coupled, Decoupled	Yes	Semi-open queueing network	Network decomposition, Regenerative processes, Markov chain analysis

using state-dependent queues. Roy and De Koster (2018) analyze the container throughput times with the QC operating in a dual-mode (both loading and unloading operations). Using a combination of open and semi-open queues, they develop an integrated stochastic model that captures the complex stochastic interactions among quayside, vehicle, and stacksides processes. The model is adopted for analyzing optimal stack layout in ALV-operated terminals. Integrated stochastic models for the seaside assume single customer class (those arriving via vessels) using SOQNs. From a theoretical point of view, there are also studies that analyze different aspects of a fork-join queueing synchronization station, which is a fundamental building block of SOQNs. Examples include studies on performance analysis (Krishnamurthy et al., 2004), scheduling and control in heavy traffic conditions (Özkan and Ward, 2019), and throughput limits (Zeng et al., 2018a,b). For a review of solution methods on semi-open queues, see Roy (2016). To analyze the vehicle type and capacity decision for inter-terminal transport vehicles, Mishra et al. (2017) develop a semi-open queueing network model with heterogeneous capacity vehicles and demonstrate the applicability with a use case of the Maasvlakte 2 terminals in the Port of Rotterdam. The semi-open queueing network model is analyzed using a free and busy period decomposition analysis. Again, they model only a single class of customers. Roy et al. (2016) carry out performance analysis of the seaside operations where AGVs are used for interterminal transport between the quay and the seaside. Dhingra et al. (2017) extend the single-stage model developed by Roy et al. (2016) to a two-stage model, where the first stage estimates the throughput parameters using the closed queueing network model. In the second stage, the throughput estimates are adopted to estimate the expected sojourn time of the vessel for both loading and unloading operations. Saini et al. (2017) develop a Markov-chain based model to estimate the crane interference delays in a twin-crane operated stack. Lee et al. (2014) use a Markov-chain based model to estimate the port capacity.

A few research papers also focus on the integrated landside operations. For example, Dhingra et al. (2018) model ET arrivals at the landside of the terminal using a two-phase Markov-modulated Poisson process and estimate the number of trucks that should be allowed in the terminal. However, the container train arrivals and the interaction with the train containers are not included in the model.

From the literature, it is evident that very few studies focus on the interaction between train and truck containers at the ASCs at the terminal landside. We address this gap by building a stochastic model that explicitly considers the service priority and the interaction between ETs and internal terminal trucks at the stacks. Research on interactions of outbound containers arriving in external trucks and trains at the yard (Storage) cranes are less understood. We particularly study the effect of priority rule and type of vehicle on the throughput time performance of external trucks and train terminal trucks. To model the congestion at the stacksides as well as the trainsides, we contribute to the literature by developing a novel multi-class semi-open queueing network model with bulk arrivals of train containers and discrete arrivals of the ETs at the stack blocks. To position this paper, Table 1 compares it with the closely relevant ones in the literature.

3. Coupled system and model description

Fig. 4 sketches a typical layout of the landside of an automated terminal (Europe Container Terminals, 2015). The storage area is divided into stack blocks, each of which has one ASC serving landside transactions (usually another ASC serves seaside transactions). The containers are stacked four or five levels high. We consider only container export operations, which includes unloading containers arriving in ETs and container trains, and storing them in the stack block. However, an analogous analysis can be made for import operations. The common notations for resources are included in Table 2. The flow of containers and the layout of the landside terminal are shown in Fig. 4.

For the coupled system, the key modeling assumptions include: 1. ETs arrive at the terminal according to a Poisson Process with rate λ (we use λ_k for ET arrivals at ASC k); 2. TTs have non-preemptive service priority over the ETs at the ASCs (typically trains

Table 2
Notations for terminal equipment.

Term	Description
ET	External Truck
TT	Terminal Truck (coupled)
LV	Lifting Vehicle (decoupled)
GC	Gantry Crane
ASC	Automated Stacking Crane

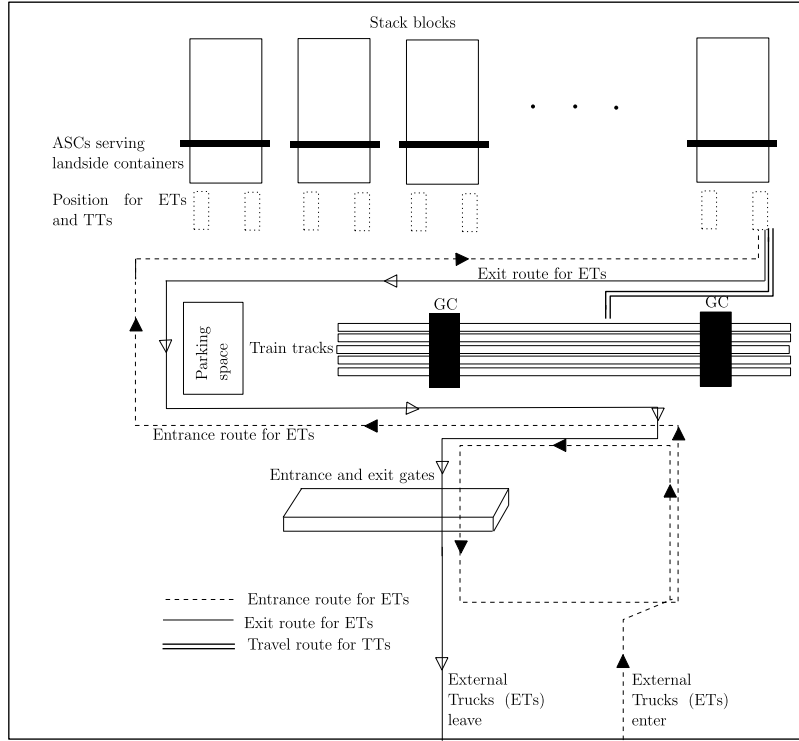


Fig. 4. Illustration of the landside terminal layout and container flow (adapted from [Dhingra et al. \(2018\)](#)).

have fixed arrival and departure schedules); 3. The GC begins its operations only when a TT is present. Likewise, ASCs begin their operations when either a TT or ET is present in the ASC buffer lane; 4. The train containers arrive with a bulk arrival process with general inter-arrival times; 5. The processing times at the GC and ASCs follow general distributions.

Upon arrival at the terminal, an ET first joins the queue at one of the terminal entry gate lanes to complete entrance formalities such as tallying truck arrival time with the appointment time slot, verifying the driver's identity, checking customs documentation, checking license and registration, and inspecting the truck for hazardous or prohibited material. In automated terminals, these formalities have already been completed at a buffer yard before the truck arrives at the gate. The ET then travels to the appropriate stack and waits for its turn. Note that an ET always requires an ASC to unload the container. After the container has been unloaded, the ET leaves the terminal. We only model single cycles of the truck (export flows via the vessel). Dual cycles of the trucks occur occasionally and can also be modeled by including additional return flows in the queuing network.

The container trains arrive at the terminal according to a renewal process with rate λ_T and squared coefficient of variation (SCV) of interarrival times, c_a^2 . A manifestation of the container train arrival process could be a deterministic schedule with fixed interarrival times (considered later for numerical experiments). Each train brings a fixed number of containers, N_{CT} . Since the trains arrive according to a renewal process, the arrival of containers on the trains form a batch renewal process. Each container on the train requests a TT before being unloaded by the rail-mounted gantry crane (GC). Once the TT arrives at the rail GC, the GC unloads the container on the TT. After being handled at the GC, the container on the train is assigned to ASC k with probability p_k . The TT now travels to the destination stack and waits in the ASC queue for its turn. We assume that the TT has non-preemptive service priority over the ET. After the container is unloaded, the TT dwells at the stackside and waits for its next job. This movement of the TT is illustrated in [Fig. 5](#). The service time at the ASCs, which includes container loading, travel, and unloading time components, is considered to have a general distribution. This matching process of a container with a TT and the movement of the TT are illustrated in [Fig. 5a](#).

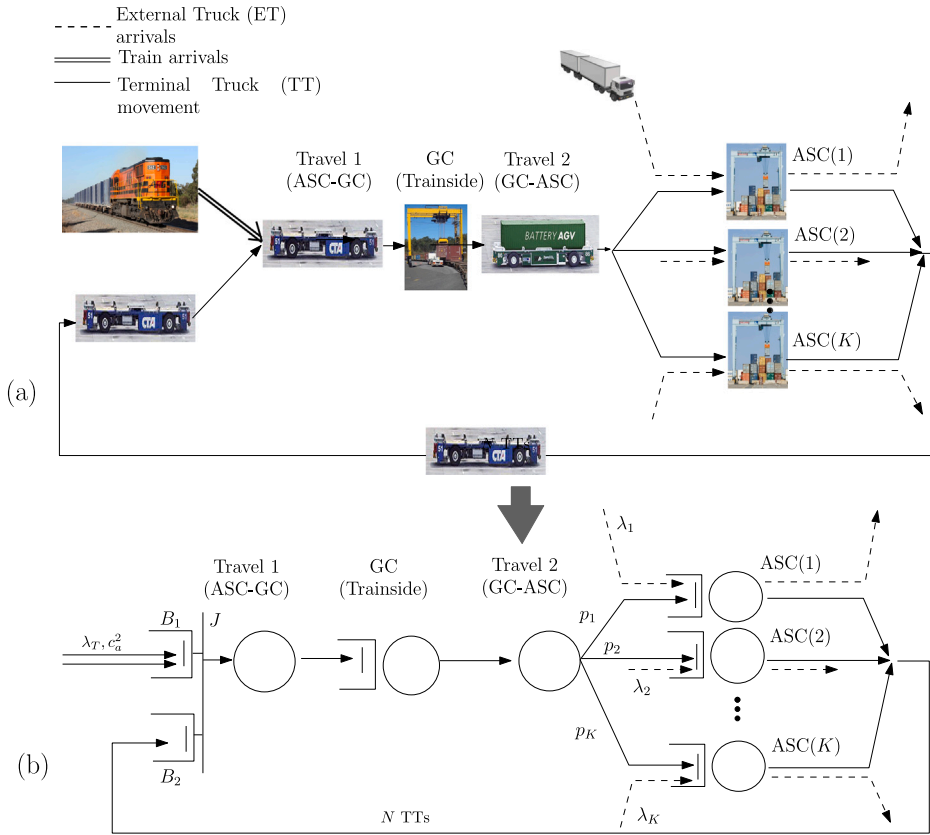


Fig. 5. (a) Illustration of the container unloading process at the landside and (b) Integrated SOQN model with ETs and train arrivals, where J denotes the synchronization station.

Fig. 5b shows the integrated model corresponding to the landside terminal processes shown in Fig. 5a. The train arrivals are modeled as a batch renewal process. The train containers wait to be assigned to a TT. TTs, which are, by assumption, coupled resources, move the containers from the trainside to the stackside. The movement of the TT from the stackside to the trainside is modeled as an Infinite Server station. The containers on a train are unloaded sequentially by a GC. A GC at the trainside is modeled as a single-server queue. The TTs queue at the GC for picking up a container. The movement of the TTs from the trainside to the stackside is also modeled as an Infinite Server station. The TTs then queue at the destination ASC, which is modeled as a single-server queue. They are routed to ASCs with uniform probability i.e., $p_k = \frac{1}{K}$. The ETs also queue at the ASC, but the TTs get non-preemptive priority over the ETs for service. The routing of the containers from the trainside to the ASC and the routing of the ETs to the ASC is random. The analysis of the coupled system is divided into three steps.

1. *Exact analysis of the system with a single ASC in isolation (Section 4)*: This step analyzes a single ASC with ET arrivals. The rest of the network (which includes the travel station, GC station) is modeled as a single state-dependent exponential server where the state-dependent rates are given. The objective of this step is to obtain the maximal network throughput when train containers are always available for unloading operations.
2. *Approximate analysis of the system with multiple ASCs (Section 5)*: This step analyzes the system with multiple ASCs with ET arrivals. Again we assume that the train containers are always available for unloading. The objective of this step is to estimate the state-dependent rates mentioned in the previous step, using an iterative approximate mean value analysis (AMVA) algorithm. Thereby, we also obtain the waiting times for the ETs at the ASCs and the network throughput of the TTs.
3. *Approximate analysis of the coupled system with synchronization station (Section 6)*: Based on the network throughput obtained in Step 2, we obtain the steady state distribution of the number of containers at the train station as seen by a container train on arrival. From this distribution, we can estimate the waiting time of the containers (arriving by train until dispatched by the TT for storage at the destination stack). In the last step, we account for the transient behavior when the train arrives in a system where some TTs are idle. Note that the waiting time estimation of the ETs in Step 2 assumes that the containers (arrived by train) are always available. Step 3 corrects for this assumption and provides better estimates of the ET waiting times. In this step, we obtain the expected unloading throughput time (time from arrival by train to storage by the ASC) by leveraging the analysis results from Step 2.

Table 3

Important quantities used in the coupled system analysis. LST denotes the Laplace–Stieltjes transform of a distribution function.

Term	Description
K	Number of ASCs in the terminal
N	Number of TTs in the terminal
N_{CT}	Number of containers on a train
D	Interarrival times of trains
λ_T, c_a^2	Rate and squared coefficient of variation (SCV) of train interarrival times with $\lambda_T^{-1} = E[D]$
λ_c, λ_{ck}	Rate of arriving train containers, i.e., $\lambda_c = \lambda_T N_{CT}$ at GC and $\lambda_{ck} = \lambda_c p_k$ at ASC (k)
S_{GC}	Service time of a TT at the GC with expectation $E[S_{GC}]$ and SCV c_{GC}^2
U_{GC}	Utilization of the GC resource
W_{GC}	Waiting time of a TT at the GC with expectation $E[W_{GC}]$
T_{GA}	Travel time of a TT from the GC to an ASC or vice versa
S_{TT}	Service time of a TT at an ASC with LST $\widehat{S_{TT}}$
S_{ET}	Service time of a ET at an ASC with LST $\widehat{S_{ET}}$
U_{TT}	Utilization of the TT
$\gamma(n)$	State dependent return rate to the ASC which we take equal to the throughput of the network consisting of travel from ASC to GC, loading of a TT at the GC and travel back to the ASC
λ_{ck}	Rate of arriving train containers at ASC k with $\lambda_{ck} = \lambda_c p_k$
λ_k	Arrival rate of ETs at ASC (k)
p_k	Probability of assigning a train container to ASC (k)
U_{ASC}	Utilization of the ASC resource
$TH_{TT,k}$	Throughput of TTs at ASC (k)
W_{TT}	Waiting time of a TT at the ASC with expectation $E[W_{TT}]$
W_{ET}	Waiting time of an ET at the ASC with expectation $E[W_{ET}]$
N_{TT}	Number of TTs at the ASC
N_{ET}	Number of ETs at the ASC
R_n	Number of TTs that return to the ASC during a service which starts with n TTs at the ASC
R_{nk}	$P(R_n = k) = \tilde{R}_{nk} - \tilde{R}_{n,k-1}$
$TH_{TT}, TH_{TT,k}$	The throughput of TTs (at ASC (k))
N_C	The difference between the number of containers that have (still) to be unloaded from the train and the number of available TTs, $N_C = -N, -N+1, \dots, 0, 1, \dots$

The queuing network model for the decoupled system is obtained by adapting the model of the coupled system. In particular, the decoupled model is an open queuing network model because the train operations and stack crane operations can be performed without the vehicle being there (see [Appendix F](#)).

4. Exact analysis of the coupled network with an ASC in isolation

We first consider a system with one ASC and two customer classes (ETs and TTs). N TTs always circulate in the system and are all served by the same ASC. After a TT leaves the ASC, it travels to the GC to pick up a container that is assumed to be always available, and then returns to the ASC. We model this trip to the GC as a state dependent exponential server with rate $\gamma(n)$, where n denotes the total number of TTs traveling to/returning back from the GC, or picking a container at the GC. At the ASC, TTs have non-preemptive priority over ETs. ETs arrive at the ASC according to a Poisson process with rate λ (see [Fig. 6a](#)).

Before we discuss the performance of the ASC in isolation given the characteristics of the state dependent server, we first concentrate on the stability of the system. Once we have the stability condition, we can focus on the joint behavior of the TTs and ETs in the system. However, we use this only as an intermediate step to find the marginal distribution of the number of TTs. Finally, we will find a way to compute the expected waiting time of the ETs.

We must also estimate the number of TTs that arrive at the ASC during a service. This obviously depends on the number of TTs that are present at the beginning of the service. Let R_n denote the number of TTs that arrive during a service when there are n TTs at the beginning of the service. In [Appendix A, Lemma 14](#), we give an explicit expression for $\tilde{R}_{nk} = P(R_n \leq k)$ for $k = 0, \dots, N - n$. Note that $\tilde{R}_{nk} = 1$ for $k = N - n, \dots, N$. For convenience, [Table 3](#) lists the key notations used in the analysis. A full definition of these and other quantities are given in the text.

Remark 1. To find the state dependent rates $\gamma(n)$, we consider the *subnetwork* consisting of the travel from the ASC to the GC (with expected travel time, $E[T_{GA}]$ and squared coefficient of variation (SCV), $c_{GA}^2 \stackrel{\text{def}}{=} \text{var}(T_{GA})/E[T_{GA}]^2$), the service at the GC (with expected duration, $E[S_{GC}]$ and SCV, c_{GC}^2) and the return to the ASC (expected time, $E[T_{GA}]$ and SCV, c_{GA}^2). Using a standard AMVA approach (see e.g., [Buzacott and Shanthikumar, 1993](#), pp 399–400), we can find the throughput of this subnetwork for n TTs, $n = 1, \dots, N$. This throughput is taken as the state dependent rate $\gamma(n)$. Assuming an exponential state dependent rate, $\gamma(n)$, the analysis of the queuing system in [Fig. 6a](#) can be performed exactly.

4.1. Stability for a network with single ASC

To find a stability condition for the network with a single ASC, we first analyze a system with the same characteristics except that we assume one ET continuously circulating in the system. This means that the ET returns immediately to the queue immediately

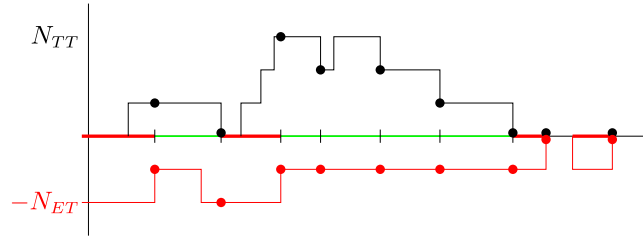


Fig. 8. Embedded process (N_{TT}, N_{ET}) just after completion of a service with multiple ETs, which are denoted by filled circles. The x axis denote the time whereas the positive y axis denotes the number of TTs and the number of ETs present at the ASC, respectively. Here the bold red lines indicate the service of ET, and green lines indicate the service of TT. This figure also illustrates that TTs have a service priority over the ETs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This regeneration cycle can be divided in service times for TTs and for ETs. By applying Wald's equation, we find that the length of the cycle, denoted by T_C , satisfies

$$E[T_C] = \left(E[N_{EOS}] - \frac{1}{1 - \widehat{S}_{ET}(\gamma(N))} \right) E[S_{TT}] + \frac{E[S_{ET}]}{(1 - \widehat{S}_{ET}(\gamma(N)))},$$

and the throughput of the ETs

$$TH_{ET} = \frac{(1 - \widehat{S}_{ET}(\gamma(N)))^{-1}}{E[T_C]} = \frac{1}{E[S_{ET}] + (P(N_{TT} = 0)^{-1} - 1)E[S_{TT}]}. \quad (3)$$

If the system in Fig. 6a is not stable, then the number of ETs will grow and there will be always a waiting ET. As soon as an ET departs, another ET is served at the ASC as long as there is no TT queued at the ASC. Hence, we need only one permanent circulating ET to determine the stability condition of the network with a single ASC. If the external arrival rate of ETs (λ) is lower than the throughput of the ET (TH_{ET}) with one permanent circulating ET (shown in Fig. 6b), then the system is always stable. On the other hand if $\lambda > TH_{ET}$, then the open system is not stable. We use this argument to develop the stability condition.

Lemma 2. *The stability condition (both necessary and sufficient) for the system with a single ASC, N TTs and ETs arriving at rate λ is*

$$\lambda E[S_{ET}] + TH_{TT} E[S_{TT}] < 1,$$

where TH_{TT} is the throughput of TTs in the corresponding closed system with one permanent ET.

Proof. Note that we can rewrite this Eq. (3) to $TH_{ET} E[S_{ET}] + TH_{TT} E[S_{TT}] = 1$ because $P(N_{TT} = 0) = TH_{ET} / (TH_{TT} + TH_{ET})$. Therefore, we can reformulate the stability condition, namely that $\lambda < TH_{ET}$ as $\lambda E[S_{ET}] + TH_{TT} E[S_{TT}] < 1$. ■

4.2. Marginal distribution of the number of TTs

Now that we know when our system is stable, we start analyzing our original system with ET arrivals (see Fig. 6(a)). In the description of our system, we keep track of the number of ETs at the ASC. Denote this number by N_{ET} . Again, we look at the embedded process at the departure epochs from the ASC (Fig. 8). Let R_{nmk} be the probability that exactly m TTs return and k ETs arrive during a service that starts with n TTs at the ASC. Note that $R_{nm} = \sum_{k=0}^{\infty} R_{nmk}$. Remark that the probability that the first service after the system empties (that is $N_{TT} = N_{ET} = 0$) is a TT, is the probability that a TT returns to the ASC earlier than an ET arrives, which equals $\gamma(N)/(\lambda + \gamma(N))$.

Theorem 3. *For the system where TTs arrive at the ASC according to state-dependent Poisson process with rate $\gamma(n)$ and ETs arrive according to a Poisson process with rate λ , the steady state probability vector for N_{TT} , $\pi \stackrel{\text{def}}{=} (P(N_{TT} = 0), \dots, P(N_{TT} = N))$, is given by $\pi = \sigma + \alpha \tau$ where σ is the steady state probability vector for the number of TTs at the ASC in case there is one permanent ET at the ASC (see Section 4.1), $\alpha = P(N_{TT} = 0, N_{ET} = 0)$ and $\Delta = \gamma(N)(R_{10} - R_{00}, \dots, R_{1,N-1} - R_{0,N-1}, R_{1N} - R_{0N})/(\lambda + \gamma(N))$.*

The proof consists of three phases: (a) we analyze the embedded process (N_{TT}, N_{ET}) where N_{TT} denotes the number of TTs at the ASC at departure epochs, (b) we find the relation between π and σ and (c) we find the constant $\alpha = P(N_{TT} = 0, N_{ET} = 0)$. Since the actual proof is quite long, it is given in Appendix C.

We use these probabilities to obtain the expected total throughput time of a TT at the ASC.

Proposition 4. *The expected total throughput time of a TT at the ASC is given by*

$$E[T_{TT}] = \frac{N\pi_N/\lambda - \sum_{n=0}^{N-1} (\pi_n R_{0N} - \pi_N R_{0n})(N-n)/\gamma(N-n)}{\sum_{n=0}^{N-1} \pi_n R_{0N} - \pi_N R_{0n}}.$$

Proof. To find the expectation of T_{TT} , we use Little's Law: $E[T_{TT}] = \frac{1}{TH_{TT}} \sum_{n=0}^N n \frac{E[T_n]}{E[T_C]}$, where $TH_{TT} = \sum_{n=0}^{N-1} \gamma(N-n)E[T_n]/E[T_C]$. Next write $\sum_{n=0}^N nE[T_n] = NE[T_C] - \sum_{n=0}^{N-1} (N-n)E[T_n]$ and use Eqs. (16) and (17) to find the expression for $E[T_{TT}]$ ■

Now we know the $E[T_{TT}]$, we can also obtain the expectation of W_{TT} , the waiting time for a TT at the ASC, which equals

$$E[W_{TT}] = E[T_{TT}] - E[S_{TT}]$$

Remark 5. In this section, we assumed a Poisson arrival process. We could extend this, following almost the same analysis, to the case of a Markov Modulated Poisson process (MMPP: see, for e.g., [Dhingra et al. \(2018\)](#)). We then not only consider the number of TTs and ETs present at the beginning of a service, but also the state s of the arrival process of the ETs. In Eq. (11) we then have to add the state of the arrival process, giving rise to a triple sum. To arrive at the transition probabilities of the embedded N_{TT} , we have to sum both over all possible number of ET arrival and over all possible states of the ET arrival process. Eqs. (12) and (13) then become

$$P(N_{TT} = n) = P(N_{TT} = 0)R_{0,n} + \sum_{\ell=1}^{n+1} P(N_{TT} = \ell)R_{\ell,n-\ell+1} \\ + P(N_{TT} = 0, N_{ET} = 0)p_{0,TT}(R_{1n} - R_{0n}), \text{ for } n = 0, \dots, N-1,$$

and

$$P(N_{TT} = N) = P(N_{TT} = 0)R_{0,N} - P(N_{TT} = 0, N_{ET} = 0)p_{0,TT}R_{0,N},$$

where $p_{0,TT}$ denotes the probability that in an empty system a TT arrives before a ET. To find this conditional probability we first find the probability $p_{0\&TT}$ that we are in an empty system and the first arriving truck is a TT. By conditioning on the state of the arrival process at the beginning of an idle period we get

$$p_{0TT} = \sum_s P(N_{TT} = 0, N_{ET} = 0, S = s)p_{TT}^{(s)},$$

where $p_{TT}^{(s)}$ is the probability that the first arriving truck after t is a TT, given that the ET's arrival process is in state s at time t and there are no TTs at the ASC. It immediately follows that $p_{0,TT} = p_{0\&TT}/P(N_{TT} = 0, N_{ET} = 0)$. The probabilities $p_{TT}^{(s)}$ can be written as $p_{TT}^{(s)} = 1 - \hat{A}^{(s)}(\gamma(N))$, where $\hat{A}^{(s)}(\cdot)$ is the LST of the time to the next arrival of a ET, starting in state s ; however, to find $P(N_{TT} = 0, N_{ET} = 0, S = s)$ a new analysis would be needed.

Next to the assumption of an MMPP for ETs, we could also assume that the service time at the ASC has a phase-type distribution. We can then model the system as a Quasi Birth–Death process (QBD) where we take as levels the number of ETs at the ASC and as phases a combination of the number of TTs at the ASC, the state of the arrival process and the state of the service time (also taking into account whether a TT or a ET is in service. This model can be analyzed using standard techniques for QBD's.

4.3. Expected waiting time of an ET

In the previous subsection, we concentrated on the number of TTs at the ASC and focused on the regeneration points where the ASC was empty. In this section, we take the same regeneration cycles, but now focus on the behavior of the ETs. We model this system as a special $M/G/1$ queue with an initial setup time for a busy period, which can be possibly zero. We remark that between the service beginnings of two subsequent ETs during a regeneration cycle, the first ET is served and, sometimes, a number of TTs are served. The time between the beginnings of two subsequent connected services of ETs in a regeneration cycle is called the modified service time.

Modified service times

The modified service time consists of two parts, the service of the ET possibly followed by a period of serving TTs. The duration of serving these TTs, call it the busy period of TTs, depends on the number of TTs that arrived during the service of the truck. Some thought reveals that the duration of a busy period of TTs starting with n TTs at the ASC is the sum of the time needed to decrease the number of TTs to $n-1$, measured from the beginning of a service of a TT, and the duration of a busy period of TTs starting with $n-1$ TTs. Let B_n denote the time needed to decrease the number of TTs from n to $n-1$, $n = 1, \dots, N$. This time itself can also be divided in different periods, namely the time to serve the TT, possibly followed by a period to serve TTs that arrived during its service. Let R_n denote the number of TTs that arrive during the service of a TT when at the start of its service n TTs are at the ASC. Note that $P(R_n = k) = P_{n,n+k-1}$ (see Eq. (1)). Conditioning on the number of TTs that return to the ASC during the first served TT, gives that

$$B_n = S_{TT} + \sum_{k=1}^{N-n} \sum_{\ell=1}^k B_{n+\ell-1} \mathbb{1}_{\{R_n=k\}} = S_{TT} + \sum_{\ell=1}^{N-n} B_{n+\ell-1} \mathbb{1}_{\{R_n \geq \ell\}},$$

where $\mathbb{1}_A$ is the indicator function of set A . Some calculus gives that

$$E[B_n] = E[S_{TT}] + \sum_{\ell=1}^{N-n} E[B_{n+\ell-1}] P(R_n \geq \ell) \quad (4)$$

and

$$\begin{aligned} E[B_n^2] &= E[S_{TT}^2] + 2 \sum_{\ell=1}^{N-n} E[S_{TT} B_{n+\ell-1} \mathbb{1}_{\{R_n \geq \ell\}}] + E\left[\left(\sum_{\ell=1}^{N-n} (B_{n+\ell-1}) \mathbb{1}_{\{R_n \geq \ell\}}\right)^2\right] \\ &= E[S_{TT}^2] + 2 \sum_{\ell=1}^{N-n} E[S_{TT} \mathbb{1}_{\{R_n \geq \ell\}}] E[B_{n+\ell-1}] + \sum_{\ell=1}^{N-n} E[B_{n+\ell-1}^2] P(R_n \geq \ell) \\ &\quad + 2 \sum_{\ell=1}^{N-n} \left(\sum_{m=1}^{\ell-1} E[B_{n+m-1}]\right) E[B_{n+\ell-1}] P(R_n \geq \ell). \end{aligned} \quad (5)$$

In [Appendix B](#), [Lemma 16](#) we give an explicit expression for $E[S_{TT} \mathbb{1}_{\{R_n \leq k\}}]$, which we can use to find $E[S_{TT} \mathbb{1}_{\{R_n \geq \ell\}}] = E[S_{TT}] - E[S_{TT} \mathbb{1}_{\{R_n \leq \ell-1\}}]$. Next remark that $B_N = S_{TT}$ and use Eqs. (4) and (5) iteratively, to find the first two moments of B_n , $n = N-1, \dots, 1$.

The modified service time S'_{ET} satisfies

$$S'_{ET} = S_{ET} + \sum_{k=1}^N \sum_{\ell=1}^k B_{\ell} \mathbb{1}_{\{R_0=k\}} = S_{ET} + \sum_{\ell=1}^N B_{\ell} \mathbb{1}_{\{R_0 \geq \ell\}},$$

where R_0 denotes the number of TTs that arrive during the service of the truck. Similar to Eqs. (4) and (5), we can find that the first two moments of S'_{ET} are given by

$$E[S'_{ET}] = E[S_{ET}] + \sum_{\ell=1}^N E[B_{\ell}] P(R_0 \geq \ell),$$

and

$$\begin{aligned} E[S_{ET}^2] &= E[S_{ET}^2] + \sum_{\ell=1}^N E[S_{ET} \mathbb{1}_{\{R_0 \geq \ell\}}] E[B_{\ell}] + \sum_{\ell=1}^N E[B_{\ell}^2] P(R_0 \geq \ell) \\ &\quad + 2 \sum_{\ell=1}^N \left(\sum_{m=1}^{\ell-1} E[B_{n+m-1}]\right) E[B_{n+\ell-1}] P(R_0 \geq \ell). \end{aligned}$$

Using the first and second moment of the modified service times, $E[S'_{ET}]$ and $E[S_{ET}^2]$, we estimate the expected total waiting time of an ET at the ASC.

Proposition 6. *The expected waiting time of an ET at the ASC is given by*

$$E[W_{ET}] = E[W_0] + \frac{\gamma(N)E[B_1]}{1 + \gamma(N)E[B_1]} \frac{E[B_1^2]}{2E[B_1]}, \quad (6)$$

where

$$E[W_0] = \frac{\lambda E[S_{ET}^2]}{2(1 - \lambda E[S'_{ET}])}. \quad (7)$$

Proof. To find the expected waiting of an ET at the ASC, we remark that the ASC starts processing either a TT or an ET after a regeneration point when the ASC is empty. We model this system, from the point of view of an ET, as an $M/G/1$ queue with an initial setup for a busy period. Note that the setup time is either zero (with probability $\lambda/(\lambda + \gamma(N))$) or B_1 (with probability $\gamma(N)/(\lambda + \gamma(N))$). In a system where the setup times is always zero, i.e., a standard $M/G/1$ queue, the expected waiting time of an ET is given by Eq. (7). For the system with setups, we see that an ET that arrives during the time the ASC is not processing an ET, can arrive in an empty system or during the setup time. By a standard argument, we let all the customers that arrive during a setup, start consecutive busy cycles in the standard queue without setups, where the waiting time is increased by B_{1R} , the remaining time of the setup, that is a busy period of TTs. The expected number of these busy cycles during a regeneration cycle is $(\lambda + \gamma(N)\lambda E[B_1])/(\lambda + \gamma(N))$. Combining these observations leads to

$$E[W_{ET}] = \frac{(\lambda E[W_0] + \gamma(N)\lambda E[B_1] (E[B_{1R}] + E[W_0]))/(\lambda + \gamma(N))}{(\lambda + \gamma(N)\lambda E[B_1])/(\lambda + \gamma(N))},$$

which can be rewritten as Eq. (6). ■

5. Approximate analysis of the coupled network with multiple ASCs

In this section, we provide an algorithm to determine the approximate performance measures of the network. We use an AMVA-like approach and relate the system with n TTs to the system with $n-1$ TTs. In the classical AMVA algorithm, we can use the queue length distribution at a station to find the expected waiting time at that station. In our system, we cannot directly relate the queue length and the waiting time at an ASC due to the possible presence of ETs. Therefore, we use the results from the previous section to compute the waiting time. Before we give results for the performance measures, we concentrate on the stability.

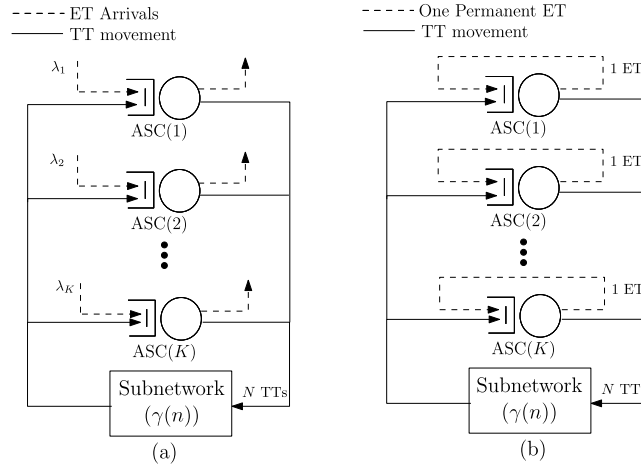


Fig. 9. (a) Queuing system with recirculating TTs, multiple ASCs, and ET arrivals, and (b) Closed system to derive stability conditions for the network with multiple ASCs.

5.1. Stability for network with multiple ASCs

Consider a closed network with multiple ASCs (Fig. 9a). We assume that there are always train containers available for pickup by the TTs. Assuming there are no ET arrivals, we estimate the throughput of the TTs ($TH_{TT,k}$) and the load at an ASC (k) ($TH_{TT,k}E[S_{TT,k}]$).

Lemma 7. For the network with multiple ASC's and ETs arriving at ASC (k) with rate λ_k ,

$$TH_{TT,k}E[S_{TT,k}] + \lambda_kE[S_{ET,k}] < 1$$

is a sufficient stability condition for ASC (k).

Proof. This condition is sufficient for stability, because the throughput of the TTs, $TH_{TT,k}$, without an ET is higher than that with an ET. ■

Next, we give a necessary stability condition. Consider a system there is always one ET present at every ASC as described in Fig. 9b. For this system, we can find $TH_{ET,k}$, the throughput of the ET at ASC (k).

Lemma 8. For the network with multiple ASC's and ETs arriving at ASC (k) with rate λ_k ,

$$TH_{ET,k} > \lambda_k \quad \forall k \in \{1, \dots, K\}$$

is a necessary stability condition for ASC (k).

Proof. This condition is necessary for stability, since the throughput of TTs, $TH_{TT,k}$, will be lower in the closed system than in the system with external arrivals where sometimes the ASCs are idle. ■

5.2. A modified AMVA algorithm for the network with multiple ASCs

Consider Fig. 6a where we have only one ASC in the network. Consider the complementary network without this ASC. We can find the throughput, $TH(n) (= \gamma(n))$, depending of the number of TTs in this complementary network by a standard AMVA algorithm. We use these state dependent throughput rates as input for the model in the previous section to find the characteristics of the ASC, especially the waiting time. Together with the waiting times at the stations in the complementary network, we can compute the throughput of the total system.

When there are more ASCs in the network, isolating one of the ASCs will not directly help us, since there are still other ASCs in the complementary network that cannot be analyzed by the AMVA directly. In this case, we iterate by isolating one ASC at a time. We first assume, that we know the waiting time at all other ASCs. Then we can find the throughput of the complementary network and can apply our findings from the previous section to compute the waiting time of the tagged ASC. We repeat this step for all ASCs. Since the waiting times at the ASCs probably change, we repeat this procedure until they no longer change. In Appendix D, we present a modified AMVA algorithm for the system with multiple ASCs based on this approach. This algorithm provides throughput of the subnetwork with n TTs, which we refer to as $\gamma(n)$. Model validation with a set of large scale experiments suggests that the percentage errors for the expected queue length performance measure estimates are about 15% (See Appendix E for details).

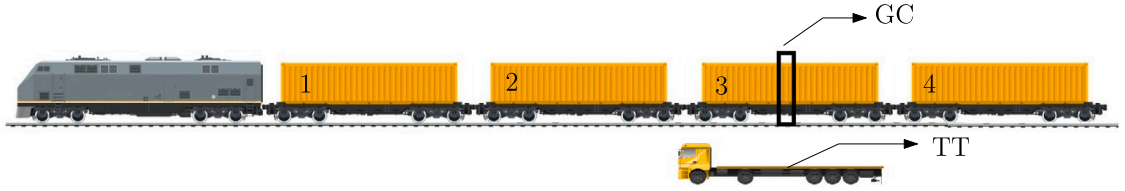


Fig. 10. Illustration of a container train unloading sequence with TTs. Note that the third container is picked up first because a TT is available for pickup from position 3, whereas the TTs for unloading containers from position 1 and 2 are en route from stackside to the trainside.

6. Coupled system performance analysis with train arrivals

In this section, we concentrate on the same system, but we now assume that the containers arrive on trains. We assume that the interarrival times of trains, denoted by D , are independent and identically distributed. The containers are unloaded sequentially depending on the availability of the TTs. Throughout this section we assume uniform handling times at the GC and deterministic travel times for the TTs. In the following, we find various performance measures for the integrated coupled system. We assume that the distribution of the handling times at an ASC is the same for TTs and ETs. Note that each container on the train (indexed one to N_{CT}) requests for a TT for movement to the stackside in FCFS sequence. However, depending on the idle position of the TT, the TT for unloading a container may not arrive in increasing order of the container index. Once a TT arrives near the container location, the GC moves to this location and unloads the container on the TT. Depending on the TT availability and the container-TT assignment sequence, the GC movement path for unloading the containers could be back and forth. Hence, the containers are not picked up from the train in FCFS sequence even though the requests for TTs by the containers are in FCFS sequence (see Fig. 10).

Stability criteria

In this section we state some necessary conditions to ensure stability of some parts of the system where trains arrive with rate λ_T each carrying N_{CT} containers and ETs arrive at ASC (k) with rate λ_k .

Lemma 9. For stability of the GC, we need $\lambda_c E[S_{GC}] < 1$, where S_{GC} is the sum of the GC travel time between any two container positions, the time for container pickup from the train, and the time for container dropoff on a TT and $\lambda_c = \lambda_T N_{CT}$ the arrival rate of train containers at the GC.

Proof. The lemma is clear by remarking that S_{GC} is the total time that it takes for the GC to handle one container. ■

Lemma 10. A necessary stability condition for the TT subsystem in the integrated coupled system network is that for every $k = 1, \dots, K$, $\widetilde{TH}_{TT,k} > \lambda_{ck}$, where $\widetilde{TH}_{TT,k}$ is the throughput of TTs at the ASC (k) with N circulating TTs, and without any ETs. A sufficient stability condition for the TT subsystem is that for every $k = 1, \dots, K$, $\underline{TH}_{TT,k} > \lambda_{ck}$, where $\underline{TH}_{TT,k}$ is the throughput of TTs at the ASC (k) with N circulating TTs, and one permanent ET at any ASC.

Proof. When we have external arrivals of ETs, the throughput of the TTs at ASC (k) denoted by TH_{TT} will satisfy $\underline{TH}_{TT,k} \leq TH_{TT,k} \leq \widetilde{TH}_{TT,k}$. ■

Lemma 11. The minimum number of TTs needed to have a stable TT-subsystem satisfies $N_{min}^{TT} > (N_{CT} \lambda_T)(S_{TT} + S_{GC} + 2T_{GA})$.

Proof. If we assume the TTs have no waiting at the ASC or the GC, we can model the system as a $GI/G/N_{min}$ multi-server queue where N_{min} the service time. Based on the queue stability condition, we have $N_{CT} \lambda_T < \frac{N_{min}^{TT}}{(S_{TT} + S_{GC} + 2T_{GA})}$ which can be rewritten to the condition in the lemma. ■

Lemma 12. To have stability of ASC (k), where ETs arrive at a rate of λ_k and train containers at rate $\lambda_{ck} = p_k \lambda_c$ where p_k is the probability that an arbitrary train container is brought to the ASC (k) it is needed that

$$U_{ASC} = \lambda_{ck} E[S_{TT}] + \lambda_k E[S_{ET}] < 1.$$

Analysis of the train station

In this section, the focus is on the train station, in particular on estimating the expected number of containers waiting to be transported by TTs and the corresponding waiting times of trains. We assume that the interarrival times of trains, denoted by D , are independent and identically distributed, that unloading takes place for one train at a time, and that the return times of TTs are exponential with rate $TH(N)$. The number of containers on a train is denoted by N_{CT} .

Just before a train arrival epoch, let N_C denote the difference between the number of containers which are to be unloaded from a train already present at the train station and the number of available TTs. Since a TT cannot be available if there are containers to

be unloaded, $N_C \geq 0$ means that there are N_C containers to be unloaded and $N_C \in \{-N, \dots, 0\}$ means that there are $-N_C$ available TTs. Due to the exponential nature of the return times of TTs, we can find the distribution of the number of returning TTs during the period between two train arrival epochs, denoted by N_{TTi} , in the same way as described in Appendix A. To simplify the analysis, we assume that the return times always occur as if all TTs are busy. In this way, we find $P(N_{TTi} = n) = \hat{D}^{(n)}(TH(N))$, where $\hat{D}^{(n)}(s)$ is the n th derivative of the LST of the interarrival times of trains. Using these probabilities, we can find the transition probabilities of the embedded Markov chain at the arrival moments of trains:

$$P(N'_C = n' | N_C = n) = \begin{cases} P(N_{TTi} = N_{CT} + n - n'), & n' > -N, \\ P(N_{TTi} \geq N_{CT} + n + N), & n' = -N, \end{cases}$$

where N'_C has the same interpretation as N_C , but then on the next train arrival epoch.

With these transition probabilities, we can now compute the stationary distribution $\pi_{TS}(n) \stackrel{\text{def}}{=} P(N_C = n)$. Once this stationary distribution is known, we need to find the expected time before a container on an arriving train is assigned to a TT. First consider the case that $-N_C = k (> 0)$ TTs are free when the train arrives. Then the $n (\leq k)$ -th container has a zero waiting time for a TT. For the $n (> k)$ -th container, the waiting time is $(n - k)/TH(N)$. Therefore, in this case, the expected time before a container is assigned to a TT, equals $\frac{(N-k)(N-k+1)}{2N_{CT}TH(N)}$. Next consider the case that $N_C = k \geq 0$. Then the average time in the system for a container on the arriving train is $\frac{1}{N_{CT}} \sum_{n=1}^{N_{CT}} (k+n) \frac{1}{TH(N)} = (k + \frac{1}{2}(N_{CT} + 1)) \frac{1}{TH(N)}$. Together these two cases provide expected waiting time for an arbitrary container on the train, $E[W_{Tr}^C]$, which is expressed as follows.

$$E[W_{Tr}^C] = \left(E[N_C | N_C \geq 0] + \frac{1}{2}(N_{CT} + 1) \right) P(N_C \geq 0) \frac{1}{TH(N)} + \frac{E[(N - N_C)(N - N_C + 1) | N_C < 0] P(N_C < 0)}{2N_{CT}TH(N)}.$$

The last part of the analysis of the train station, is to find the time that all TTs are free. This idle time has to be analyzed to find the characteristics of the arrival process of TTs to the ASCs, which is needed in the next section. During the idle time, the ETs at the ASCs do not have any interactions with the TTs. Now consider a busy period at the train station. During this period there is only one train that arrives when all the TTs are free, so on average $1/\pi_{TS}(-N)$ trains are handled and the expected length of the busy period is $E[N_{CT}] / (\pi_{TS}(-N)TH(N))$, while the fraction of time the TTs are busy equals $E[N_{CT}] / (TH(N)E[D])$. The expected length of an idle period therefore equals $(E[D]TH(N) - E[N_{CT}]) / (\pi_{TS}(0)TH(N))$.

Analysis of the waiting time at an ASC

To analyze the waiting time at an ASC, we need a different approach for TTs and ETs. For the expected waiting time of a TT at an ASC, denoted by W_{TT} , we use the model where we assume that there are always train containers to be handled (see Section 5). This same model can be used to find the utilization of the TTs by calculating the total time needed to handle the number of containers on a train and by dividing this time by the interarrival times of trains. This provides an estimate of TT utilization (see Eq. (8)).

$$U_{TT} = (E[W_{TT}] + E[T_{GC}] + 2E[T_{GA}])N_{CT}/NE[D] \quad (8)$$

where $E[T_{GC}]$ is the expected throughput time of a TT at the GC. For the expected waiting time for an ET at the ASC, denoted by W_{ET} , we first consider the ASC as a $GI/GI/1$ queue. Denote the squared coefficient of variation (SCV) for the ET interarrival times and the ASC service times by c_a^2 and c_s^2 respectively and assume, for the moment, that both SCVs are known. Then the well known two moment approximation $E[W_{ASC}] = \frac{c_a^2 + c_s^2}{2} \frac{U_{ASC}}{1 - U_{ASC}} E[S_{ASC}]$ can be used to approximate the expected waiting time of an arbitrary truck. By assuming that the handling times of TTs and ETs have the same expectation, the total number of trucks at the ASC with the priority of TTs over ETs, is the same as the number of trucks in the $GI/GI/1$ queue. We then find that for the priority queue

$$E[W_{ET}]p_{ET} + E[W_{TT}](1 - p_{ET}) = E[W_{ASC}], \quad (9)$$

where $p_{ET} = \lambda_k / (\lambda_k + \lambda_{ck})$. It remains to find c_a^2 and c_s^2 . The SCV of the service times is easily found to be $c_s^2 = E[S_{ASC}^2] / (E[S_{ASC}])^2 - 1$. The arrival process is comprised of two streams: (1) a Poisson arrival stream of ETs, and (2) a general process of arriving TTs with a certain squared coefficient of variation, $c_{a,TT}^2$, which can be determined by considering the departure process at the GC and the routing probabilities. Using this observation, we can approximate c_a^2 . Together with the moment approximation for the expected waiting time in the $GI/G/1$ queue, this gives $E[W_{ASC}]$, the approximated expected waiting times for any truck in the system without priority for the TTs and, by using Eq. (9), an approximation for $E[W_{ET}]$. Refer to Appendix F for the analysis of the decoupled system with train arrivals.

7. Model validation and insights

We first validate our model with large scale instances. The test data is obtained from the Port of Rotterdam, APM terminals. We consider the terminal with 14 ASCs, two levels of train arrivals (8 and 24 per day), seven levels of external truck arrivals (from 86 trucks per hour to 216 trucks per hour), and two levels of the number of terminal trucks (6 and 10), leading to 28 instances in total. The speed for both coupled and decoupled vehicles are set at 6 m/s. The speed for an ASC is set at 3 m/s with additional 20 s duration for picking up and 20 s duration for setting down tasks. We consider one GC handling train containers. The speed of a GC is set at 2.5 m/s with additional 12 s duration for picking up and 12 s duration for setting down containers. Each train

Table 4
Data for experiments.

Stackside	Vehicle transport	Trainside
14 stacks, six buffer lanes	Area of 70 m × 620 m	585 m long (40 containers)
Each stack has 10 rows, 32 bays, five tiers	6, 10 TTs or LVs	1 GC
Velocity 3 m/s	Velocity 6 m/s	Velocity 2.5 m/s

Table 5
Summary statistics of the percentage errors for the coupled system, $\frac{A-S}{S} \times 100\%$.

Statistic	U_{TT}	U_{GC}	U_{ASC}	$E[W_{TT}]$	$E[W_{ET}]$	$E[W_{GC}]$	$E[W_{Tr}^C]$
Maximum	3.22%	-2.53%	1.81%	8.38%	10.77%	3.39%	13.55%
Minimum	-1.52%	-14.39%	-1.63%	2.00%	-7.47%	-15.54%	-44.40%
Median	-0.41%	-9.77%	0.41%	6.73%	0.62%	-13.67%	6.70%
Average	0.45%	-9.56%	0.34%	6.39%	0.93%	-9.14%	2.63%

Table 6
Summary statistics of the percentage errors for the decoupled system, $\frac{A-S}{S} \times 100\%$.

Statistic	U_{LV}	U_{GC}	U_{ASC}	$E[W_{TT}]$	$E[W_{ET}]$	$E[W_{GC}]$
Maximum	0.12%	-0.79%	1.75%	26.72%	8.89%	0.95%
Minimum	-0.14%	-1.02%	-0.52%	-5.03%	5.19%	0.90%
Median	0.09%	-0.90%	0.46%	4.50%	6.95%	0.92%
Average	0.04%	-0.90%	0.50%	5.95%	7.03%	0.92%

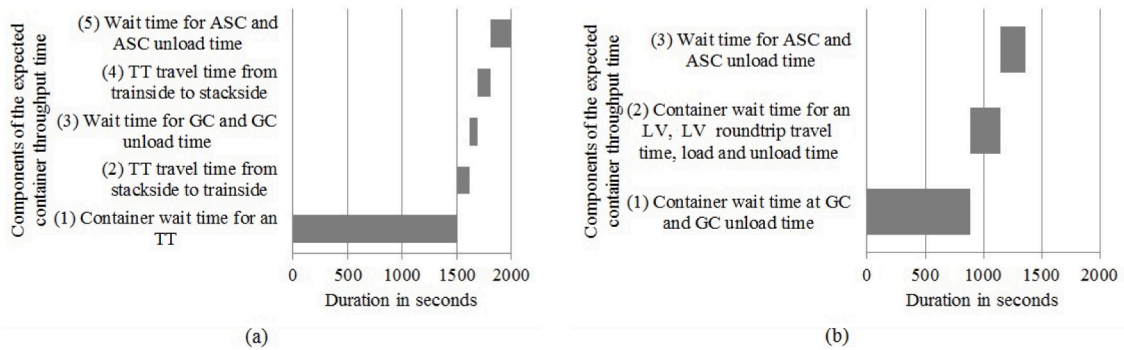


Fig. 11. Components of expected container throughput time, (a) coupled system and (b) decoupled system for 6 TTs/ LVs, 8 trains per day, and 86.4 ETs arrive per hour.

brings 40 containers to the terminal. The two versions of the simulation model for the landside container terminal (coupled and decoupled system) are developed using AutoMod simulation software (www.automod.com). In the simulation, the ASCs and the GC are modeled using a bridge crane system whereas the transport vehicles and their travel paths are modeled with the path mover system. This model is close to reality because the physical configuration of the vehicle paths and real operation of the GC and ASCs are modeled. Each scenario is run for 15 replications and 95% confidence intervals for the performance measures are obtained. The replication length is set to 20 days. The data used in the analysis is included in Table 4. The stack layout data is obtained from Roy et al. (2020). The train arrival data is obtained from Rotterdam terminals. Other data is obtained from terminal experts.

We first discuss the comparison of the performance measures obtained from the analytical model and simulation (refer Tables 5 and 6). In the test cases for the coupled system, the utilization of the TTs, ASCs, and GC range between 20%–100%, 35%–93%, and 16%–55%, respectively. For the same parameters, the resource utilization of LVs, ASCs, and GC in the decoupled system, range between 9%–43%, 35%–93%, and 11%–33%, respectively. For all performance measures, the average errors, reported as $\frac{(A-S)}{S}$ where A and S are performance measure estimates obtained from analytical and simulation models respectively, are less than 10%.

7.1. Container waiting time distribution at trainside: Comparison of coupled vs decoupled system

Using the analytical models, we illustrate the container throughput time of a coupled and decoupled system using a stacked bar chart (see Fig. 11). We find that the average waiting time of the containers on the train in the coupled system is almost twice as high as in the decoupled system. This waiting time is the most significant component of the total container throughput time (75% in the coupled system and 65% in the decoupled system). Better scheduling and dwell point selection of the coupled vehicles may not reduce this waiting time sufficiently because the train containers arrive in batches. Hence, all vehicles are busy at the same time.

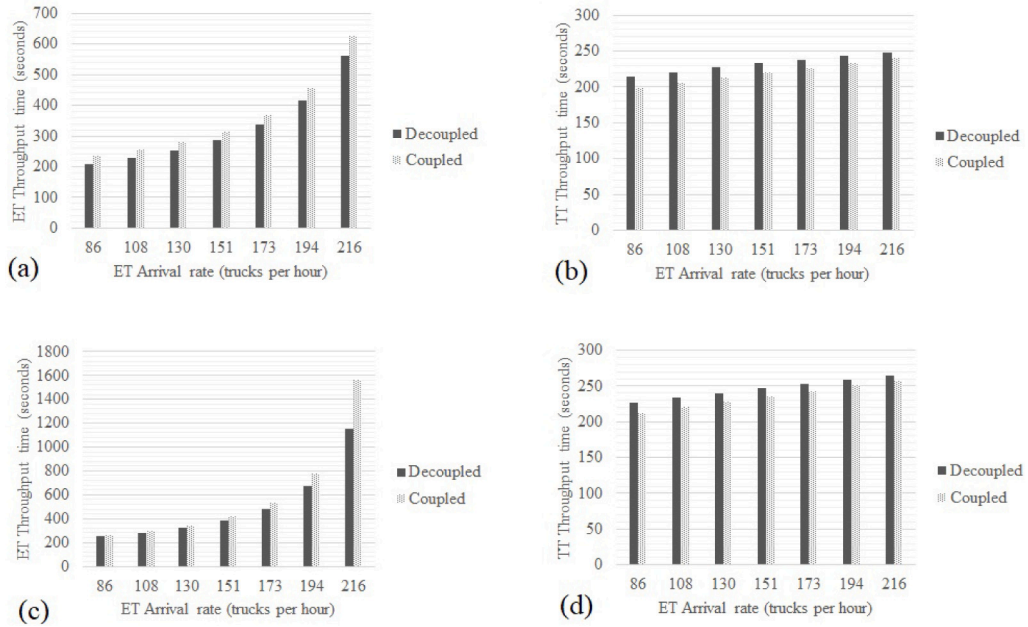


Fig. 12. (a,b) Expected throughput time for the ET and TT containers at the ASC, respectively with 6 vehicles and 8 trains per day, and (c,d) Expected throughput time for the ET and TT containers at the ASC, respectively with 10 vehicles and 24 trains per day.

In this numerical experiment, we observe that the waiting time for a TT is almost double for a container in the coupled system than the decoupled system. Note that for the comparison, we maintain the same number of vehicles in both cases. In the coupled system, the GC can only unload the container from the train once a TT is available to pick up. This dependence increases the waiting time for the containers. On the other hand, the container can be immediately unloaded from the train when a GC is available. The container waits at the train buffer lane for the LV to pick up. Hence, the decoupled nature of the processes reduces the container waiting time for unloading. The distribution of the container waiting times on the train for coupled systems also shows higher variability of the average container waiting times ($CV=0.28$) compared to the decoupled system ($CV=0.06$). Especially, the distribution of the waiting times has a long tail in the coupled system, which occurs at very high TT utilization. Note that when trains arrive in the decoupled system, all containers are unloaded with almost deterministic GC handling times. Hence, the average container waiting times in the decoupled case have a negligible variance.

7.2. Effect of resource flexibility on container waiting times at ASC: Comparison of coupled vs decoupled system

Using analytical models, we analyze the throughput times of the ETs for coupled and decoupled internal transport with varying external truck arrival rates, 86 to 216 trucks per hour among all ASCs. We have two main observations: (a) Comparing Fig. 12(a)–(d), TTs always have the same or shorter average throughput time than ETs for the same ET arrival rate for both coupled and decoupled systems. This phenomenon is to be expected as TTs get non-preemptive priority over ETs at the ASCs. The effect becomes even more pronounced for higher ET arrival rates. (b) From Fig. 12 (a,c), we observe that the throughput time of the external truck containers is lower with decoupled resources in comparison to throughput time obtained with coupled resources. For the ETs, it is likely that containers find the ASC in the decoupled system more empty upon arrival in comparison to the coupled system. However, from Fig. 12 (b,d), the TT throughput time at the ASC is higher with decoupled resources in comparison to train container throughput time with coupled resources. Since the TTs in the coupled system queue at the single GC to pick up the train container, the variability in TT return times to the ASC is low; however, with decoupled vehicles, the LVs pick up container from the train side and dropoff the train containers at the ASC without resource coordination. Hence, variability in vehicle return times to the ASC is high and more train containers queue at the ASC in decoupled system that results in higher container throughput time. Fig. 12(a,b) is based on few train arrivals per day, whereas Fig. 12(c,d) show results for a large number of train arrivals per day. Figs. 12(b,d) show that the train container throughput times at the ASC are not affected much at varying train arrival rates. This effect is observed because constant number of internal vehicles move the train containers. However, ETs experience longer waits at the ASC at high train arrival rates i.e., when the train arrivals per day increase from 8 to 24 per day, ET containers waiting times at the ASC almost doubles. For both low and high train arrival rates, the ET container customers realize more throughput time benefits with the decoupled transport resources (up to 12%). However, decoupled vehicles marginally increase the train container handling delays at shared stack cranes (up to 6%).

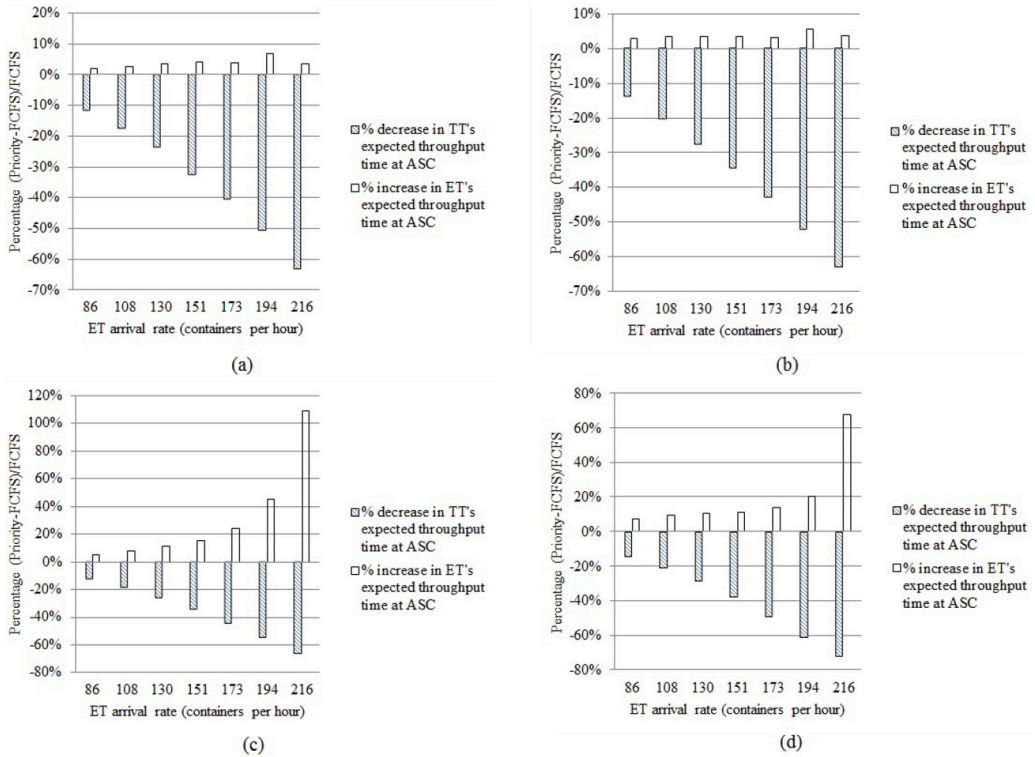


Fig. 13. Effect of service priority on TT and ET containers at ASC, (a) $N=6, \lambda_T=8/\text{day}$, (b) $N=10, \lambda_T=8/\text{day}$, (c) $N=6, \lambda_T=24/\text{day}$ (b) $N=10, \lambda_T=24/\text{day}$.

7.3. Effect of resource priority on container waiting times at ASC: Comparison of coupled vs decoupled system

We simulate the ASC performance for two situations to study the effect of priority on container waiting times. First, with priority of TT containers over ET containers at the ASC and second, with FCFS processing of the containers at the ASC. For the coupled case, we consider four scenarios where both the number of TTs and the train arrival rates are varied at two levels (see Fig. 13). For all four scenarios, we increase the aggregate ET arrival rate from 86 to 216 trucks per hour. For cases with low train arrival rate (8/day), we observe that priority for TT containers leads to a significant reduction in the expected throughput time at the ASCs in comparison to the FCFS scheduling policy (12%–63%) and a small increase in the expected throughput time for the ETs at the ASCs (2%–7%). For high train arrival rates (24/day), the ET throughput times are affected more. We observe that the priority for TT containers still significantly reduced the expected TT throughput time (12%–72%) at the ASCs; however, now the increase in the expected throughput time for the ETs at the ASCs is much higher (5%–109%). In particular, a high increase in ET throughput times occurs when N is low and the ET and train arrival rates are very high.

We also do additional analysis on how the system performance compare when the ET arrival rates follow a Markov Modulated Poisson process (see Appendix G). We also do additional cost–benefit analysis by finding the minimum number of vehicles in coupled and decoupled system (see Appendix H). To find the minimum number of TTs in a system with 4 ASCs, all with a mean service time of 157 s, we consider a system with a permanent external truck present at every ASC. The travel times from GC to ASC (116 s) and the service time at the GC (42 s). This is a good representation of a highly loaded system, since the TTs have priority over the external trucks. For the coupled system, we compute the throughput of the TTs through the system that have always a permanent truck for every ASC stations and V circulating TTs. The throughput should at least match the train arrival rate with bulk containers. For the decoupled system, we account for the vehicle travel times along with LV loading and unloading time.

The key observations do not change: (1) ET throughput times are greater than TT throughput times, for all arrival rates, (2) TT throughput times are fairly insensitive to arrival rates (for both coupled and decoupled systems), (3) Prioritizing the train containers do not affect the ET throughput times with decoupled vehicles, and (4) ET throughput time is lower with decoupled vehicles than coupled vehicles and TT throughput time is lower with coupled vehicles than decoupled vehicles.

8. Conclusions and future work

We present a stochastic model for analyzing landside processes at a multi-modal container terminal. In particular, we model the effect on container throughput times of prioritizing train containers that arrive in bulk, over truck containers at the stack cranes and the use of coupled and decoupled container transport vehicles. We develop a stylized semi-open queuing network with

external arrivals at a synchronization buffer (representing the train gantry crane with bulk arrivals) as well as at an internal station (representing the stack cranes). The special features make the model hard to analyze. While a special case can be solved exactly, approximate solution methods are developed using regenerative process analysis for the general case.

We show that decoupled resources reduce ET throughput times but increase them for LVs. Hard-coupled resources can enforce coordination and minimize congestion at the ASC, but can increase train sojourn time because the containers are only removed from the train if the TTs are available. A decoupled system can decrease the sojourn time of the trains but requires more buffer space in front of the container train. So the trade-offs between buffer space costs and additional costs of decoupled resources need to be analyzed. Decoupling resources would lead to a reduction of LVs and could increase investment feasibility. Moreover, decoupled vehicles are expensive. A two-stage decoupled approach, such as a truck with a reach stacker, could be applied to reduce the degree of coupling.

One straightforward application of our model is sizing the number of resources (number of coupled vs. decoupled resources) based on the technology choice for a given throughput level. Also, our model can be used to find the effect of vehicle path topology on reducing the travel times and number of vehicles. It is also useful for examining the implications of dedicated vs. pooled stacks (stacks dedicated to train container movement and external truck container movement). Researchers may also extend our model to understand the implications of dual command cycles in increasing system throughput. Further, our approximate analysis approach can be adopted for performance analysis of other transportation systems with large number of shared resources. Partition the train side into zones with one gantry crane allocated to each zone. Another research could be to analyze the coupled system performance by partitioning train length into multiple zones and allocate a GC to each zone that can work independently without blocking. This scenario can reduce the waiting at the GC. Our proposed model can be adapted for simultaneous loading and unloading at both train side and stack side with new flows, new processing times at the ASCs and GC, and can form an interesting future study.

CRediT authorship contribution statement

Debjit Roy: Conceptualization, Methodology, Software, Writing – review & editing, Data curation, Result interpretation. **Jan-Kees van Ommeren:** Conceptualization, Methodology, Software, Writing – review & editing, Result interpretation. **René de Koster:** Model discussion, Result interpretation, Writing – review & editing, Data curation. **Amir Gharehgozli:** Model discussion, Result interpretation, Writing – review & editing, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Arrival process of TTs at the ASC

In this part, we focus on the number of TTs that return to the ASC during a service time that starts with n TTs, $n = 0, \dots, N$, present at the beginning of the service. We start with a general lemma about the sum of exponentially distributed random variables.

Lemma 13. Let $X = \sum_{i=1}^n X_i$, where X_i , $i = 1, \dots, n$ are independent exponentially distributed random variables with $E[X_i] = 1/\beta_i$. Define $S = \{\beta_1, \dots, \beta_n\} = \{\tilde{\beta}_1, \dots, \tilde{\beta}_m\}$ with $\tilde{\beta}_1 < \dots < \tilde{\beta}_m$ and let $m_i = \sum_{\ell=1}^n \mathbb{1}_{\{\beta_\ell = \tilde{\beta}_i\}}$. Then the probability density function (pdf) of X for $t \geq 0$ is given by

$$f_X(t) = \sum_{i=1}^m \sum_{\ell=0}^{m_i-1} p_{i,m_i-\ell-1} \frac{t^\ell}{\ell!} e^{-\tilde{\beta}_i t},$$

where the $p_{i\ell}$ satisfy

$$\left(\frac{d}{ds}\right)^k \left(\prod_{j=1, j \neq i}^m (\tilde{\beta}_j + s)^{m_j} \sum_{\ell=0}^{m_i-1} p_{i\ell} (\tilde{\beta}_i + s)^\ell \right)_{s=-\tilde{\beta}_i} = \mathbb{1}_{\{k=0\}} \prod_{i=1}^n \beta_i^{m_i},$$

for all $k = 0, \dots, m_i - 1$ and $i = 1, \dots, m$.

During a service at the ASC, the number of TTs can only increase when TTs return. Denote the number of returning TTs during a service that starts with $N_{TT} = n$ by R_n . Define $\tilde{R}_{nk} \stackrel{\text{def}}{=} P(R_n \leq k)$ and set $\tilde{R}_{n,-1} \stackrel{\text{def}}{=} 0$ and $\tilde{R}_{n,N-n} \stackrel{\text{def}}{=} 1$. By assumption, the state dependent server that generates the returning TTs, has exponential service times with rate $\gamma(n)$, $n = 0, \dots, N$ with $\gamma(0) = 0$. Let $S_{nk} = \{\gamma(N-n-k), \gamma(N-n-k+1), \dots, \gamma(N-n)\}$; since rates at different states can be the same, we write $S_{nk} = \{\gamma_{nk1}, \dots, \gamma_{nkm_{nk}}\}$ with $\gamma_{nk1} < \dots < \gamma_{nkm_{nk}}$ for $k = 0, \dots, N-n-1$ and $n = 0, \dots, N-1$ and let $m_{nki} = \sum_{\ell=0}^k \mathbb{1}_{\{\gamma(N-n-\ell) = \gamma_{nki}\}}$ denote the multiplicity of the rate γ_{nki} for $i = 1, \dots, m_{nk}$.

Lemma 14.

$$\tilde{R}_{nk} = \sum_{i=1}^{m_{nk}} \sum_{j=0}^{m_{nki}-1} \frac{(-1)^j \hat{S}^{(j)}(\gamma_{nki})}{j!} \sum_{\ell=0}^{m_{nki}-1-j} p_{nki\ell} \gamma_{nki}^{\ell+j-m_{nki}} \quad (10)$$

where the constants $p_{nki\ell}$ are defined analogously as in Lemma 13 and $\hat{S} = \widehat{S_{TT}}$ if $n > 0$ and $\hat{S} = \widehat{S_{ET}}$ if $n = 0$;

Proof. We assume that n TTs are at the ASC (and therefore $N - n$ TTs are at the state dependent server). Then it is easily seen that $P(R_n \leq k) = P(\sum_{\ell=0}^k X_{N-n-\ell} > S)$ for $k = 0, \dots, N - n - 1$ and $P(R_n \leq k) = 1$ for $k = N - n, \dots, N$ where X_ℓ are independent exponentially distributed random variables with rate $\gamma(\ell)$ for $\ell = 1, \dots, n$. Following Lemma 13, the pdf of $\sum_{\ell=0}^k X_{N-n-\ell}$ is given by

$$f_{nk}(t) = \sum_{i=1}^{m_{nk}} \sum_{\ell=0}^{m_{nk}-1} p_{nki, m_{nk}-1-\ell} \frac{t^\ell}{\ell!} \exp(-\gamma_{nki} t).$$

For $k = 0, \dots, N - n - 1$, condition on the length of the service time S to find

$$\begin{aligned} \tilde{R}_{nk} &= \int_0^\infty \int_s^\infty f_{nk}(t) dt dS(s) \\ &= \sum_{i=1}^{m_{nk}} \sum_{j=0}^{m_{nk}-1} \frac{(-1)^j \hat{S}^{(j)}(\gamma_{nki})}{j!} \sum_{\ell=0}^{m_{nk}-1-j} p_{nki} \gamma_{nki}^{\ell+j-m_{nk}}. \quad \blacksquare \end{aligned}$$

Corollary 15. When the return rates increase when more TTs are traveling, i.e., $\gamma(1) < \dots < \gamma(N)$ the constants p_{nki} are given by

$$p_{nki} = \gamma(N - k + i) \prod_{j=1, j \neq i}^n \frac{\gamma(N - k + j)}{\gamma(N - k + j) - \gamma(N - k + i)}.$$

Appendix B. The service times of a TT at the ASC

In this appendix, we study the moments of the service time of a TT at the ASC which starts with n TTs present and during which at most k TTs arrive at this ASC.

Lemma 16. For $n = 1, \dots, N - 1$ and $k = 0, \dots, N - n$

$$E \left[S_{TT}^p \mathbb{1}_{\{R_n \leq k\}} \right] = \sum_{i=1}^{m_{nk}} \sum_{j=0}^{m_{nk}-1} \frac{(-1)^{j+p} \hat{S}_{TT}^{(j+p)}(\gamma_{nki})}{j!} \sum_{\ell=0}^{m_{nk}-1-j} p_{nki} \gamma_{nki}^{\ell+j-m_{nk}}$$

Proof. In a similar way as we derived R_{nk} in Eq. (10), we find for $k = 0, \dots, N - 1$ that

$$\begin{aligned} E \left[S_{TT}^p \mathbb{1}_{\{R_n \leq k\}} \right] &= \int_0^\infty s^p \int_s^\infty f_{nk}(t) dt dS_{TT}(s) \\ &= \sum_{i=1}^{m_{nk}} \sum_{\ell=0}^{m_{nk}-1} p_{nki, m_{nk}-1-\ell} \int_0^\infty s^p \int_s^\infty \frac{t^\ell}{\ell!} \exp(-\gamma_{nki} t) dt dS_{TT}(s) \\ &= \sum_{i=1}^{m_{nk}} \sum_{\ell=0}^{m_{nk}-1} p_{nki, m_{nk}-1-\ell} \int_0^\infty \sum_{j=0}^\ell \frac{s^{j+p}}{j!} \frac{\exp(-\gamma_{nki} s)}{\gamma_{nki}^{\ell+1-j}} dS_{TT}(s) \\ &= \sum_{i=1}^{m_{nk}} \sum_{j=0}^{m_{nk}-1} \frac{(-1)^{j+p} \hat{S}_{TT}^{(j+p)}(\gamma_{nki})}{j!} \sum_{\ell=j}^{m_{nk}-1} \frac{p_{nki, m_{nk}-1-\ell}}{\gamma_{nki}^{\ell+1-j}} \\ &= \sum_{i=1}^{m_{nk}} \sum_{j=0}^{m_{nk}-1} \frac{(-1)^{j+p} \hat{S}_{TT}^{(j+p)}(\gamma_{nki})}{j!} \sum_{\ell=0}^{m_{nk}-1-j} p_{nki} \gamma_{nki}^{\ell+j-m_{nk}} \quad \blacksquare \end{aligned}$$

Appendix C. Proof of Theorem 3

In this appendix, we give the proof of Theorem 3. For clarity, we first repeat the theorem.

Theorem 3. For the system where TTs arrive at the ASC according to state-dependent Poisson process with rate $\gamma(n)$ and ETs arrive according to a Poisson process with rate λ , the steady state probability vector for N_{TT} , $\pi \stackrel{\text{def}}{=} (P(N_{TT} = 0), \dots, P(N_{TT} = N))$, is given by $\pi = \sigma + \alpha \tau$ where σ is the steady state probability vector for the number of TTs at the ASC in case there is one permanent ET at the ASC (see Section 4.1), $\alpha = P(N_{TT} = 0, N_{ET} = 0)$ and $\Delta = \gamma(N)(R_{10} - R_{00}, \dots, R_{1, N-1} - R_{0, N-1}, R_{1N} - R_{0N})/(\lambda + \gamma(N))$.

Proof. The proof consists of three phases: (a) we analyze the embedded process (N_{TT}, N_{ET}) where N_{TT} denotes the number of TTs at the ASC at departure epochs, (b) we find the relation between π and σ and (c) we find the constant $\alpha = P(N_{TT} = 0, N_{ET} = 0)$.

First consider Part (a), the embedded process (N_{TT}, N_{ET}) . For this process, we can write down the balance equations for $k = 0, 1, \dots$:

$$\begin{aligned} P(N_{TT} = n, N_{ET} = k) = & \sum_{i=0}^k \sum_{\ell=1}^{n+1} P(N_{TT} = \ell, N_{ET} = i) R_{\ell, n-\ell+1, k-i} \\ & + \sum_{i=1}^{k+1} P(N_{TT} = 0, N_{ET} = i) R_{0, n, k+1-i} \\ & + P(N_{TT} = 0, N_{ET} = 0) \left(\frac{\lambda}{\lambda + \gamma(N)} R_{0nk} + \frac{\gamma(N)}{\lambda + \gamma(N)} R_{lnk} \right), \end{aligned} \quad (11)$$

for $n = 0, \dots, N-1$, and

$$\begin{aligned} P(N_{TT} = N, N_{ET} = k) = & \sum_{i=1}^{k+1} P(N_{TT} = 0, N_{ET} = i) R_{0, N, k+1-i} \\ & + P(N_{TT} = 0, N_{ET} = 0) \frac{\lambda}{\lambda + \gamma(N)} R_{0Nk}, \end{aligned}$$

To find the last balance equation, note that at the end of a service $N_{TT} = N$ is only possible if no TT is being served so all TTs were not at the ASC at the beginning of this service and they all returned.

Next, consider Part (b), relating π and σ . Summing these balance equations over k , i.e., all possible outcomes of N_{ET} , and interchanging the order of summation, eventually leads to

$$\begin{aligned} P(N_{TT} = n) = & \sum_{\ell=1}^{n+1} P(N_{TT} = \ell) R_{\ell, n-\ell+1} \\ & + P(N_{TT} = 0, N_{ET} > 0) R_{0, n} \\ & + P(N_{TT} = 0, N_{ET} = 0) \left(\frac{\lambda}{\lambda + \gamma(N)} R_{0n} + \frac{\gamma(N)}{\lambda + \gamma(N)} R_{ln} \right), \end{aligned}$$

for $n = 0, \dots, N-1$, and

$$P(N_{TT} = N) = P(N_{TT} = 0, N_{ET} > 0) R_{0, N} + P(N_{TT} = 0, N_{ET} = 0) \frac{\lambda}{\lambda + \gamma(N)} R_{0N},$$

Finally, by writing $P(N_{TT} = 0) = P(N_{TT} = 0, N_{ET} > 0) + P(N_{TT} = 0, N_{ET} = 0)$, we find

$$\begin{aligned} P(N_{TT} = n) = & P(N_{TT} = 0) R_{0, n} + \sum_{\ell=1}^{n+1} P(N_{TT} = \ell) R_{\ell, n-\ell+1} \\ & + P(N_{TT} = 0, N_{ET} = 0) \frac{\gamma(N)}{\lambda + \gamma(N)} (R_{ln} - R_{0n}), \text{ for } n = 0, \dots, N-1, \end{aligned} \quad (12)$$

and

$$P(N_{TT} = N) = P(N_{TT} = 0) R_{0N} - P(N_{TT} = 0, N_{ET} = 0) \frac{\gamma(N)}{\lambda + \gamma(N)} R_{0N}. \quad (13)$$

Since the (N_{TT}, N_{ET}) -process at the embedded point is an aperiodic Markov chain, it has a steady state distribution, and so the $\pi \stackrel{\text{def}}{=} (P(N_{TT} = 0), \dots, P(N_{TT} = N))$ also exists.

We can write the set of Eqs. (12) and (13) in a slightly more abstract way as $\pi = \pi P + \alpha \Delta$ where P is the probability matrix defined in Eq. (1).

Look at the difference of two steady state probability vectors, $\delta = \pi - \sigma$ where σ is the steady state distribution of the number of TTs at the ASC in the modified system with one permanent ET, see Eq. (2). Then $\sum_{\ell=0}^N \delta_{\ell} = \sum_{\ell=0}^N \pi_{\ell} - \sigma_{\ell} = \sum_{\ell=0}^N \pi_{\ell} - \sum_{\ell=0}^N \sigma_{\ell} = 1 - 1 = 0$ and $\delta = \delta P + \alpha \Delta$. So, the vector $\tau = \delta/\alpha$ satisfies $\tau = \tau P + \Delta$. To find τ , we can solve $\tau \tilde{P} = \tilde{\Delta}$ where $\tilde{\Delta} = \Delta - \gamma(N)(R_{1\ell} - R_{0\ell})e_{\ell}/(\lambda + \gamma(N))$, cf. Eq. (2).

Since

$$\pi = \sigma + \alpha \tau, \quad (14)$$

effectively, we only have to find α to get π .

Finally, consider Part (c) finding α : We again use the theory of regenerative processes. Define a regeneration point as an epoch in which the ASC becomes empty, that is both $N_{TT} = 0$ and $N_{ET} = 0$. This regeneration point is for both the continuous time process and the embedded process. We denote the length of the continuous time regeneration cycle by T_C . By an up-down crossing argument, the number of TTs during a cycle that leave n TTs at the ASC is equal to the number of TTs that see n TTs at the ASC upon arrival. Let T_n be the total time during a cycle that n TTs are at the ASC (and therefore $N - n$ TTs are traveling). Then the expected number of TTs that see n TTs at the ASC during a cycle equals $\gamma(N - n)E[T_n]$. The expected number of ETs that leave n TTs at the ASC equals $\lambda E[T_C] R_{0n}$. Since there is only one truck that ends a cycle, it follows from the theory of regenerative processes that the expected total number of trucks (TTs and ETs) that are handled during this cycle equals $1/\alpha$. Therefore

$$\pi_n = \frac{\gamma(N - n)E[T_n] + \lambda E[T_C] R_{0n}}{1/\alpha}. \quad (15)$$

Next take $n = N$ to find that

$$\lambda E[T_C] = \frac{\pi_N}{\alpha R_{0N}}, \quad (16)$$

since $\gamma(0) = 0$. With Eq. (15) this gives that

$$\gamma(N - n)E[T_n] = \frac{\pi_n R_{0N} - \pi_N R_{0n}}{\alpha R_{0N}}, \quad (17)$$

for $n = 0, \dots, N - 1$. Since a cycle starts with an idle period, followed by (multiple) periods where ETs or TTs are served, the expected cycle length $E[T_C]$ satisfies

$$E[T_C] = \frac{1}{\lambda + \gamma(N)} + \lambda E[T_C] E[S_{ET}] + \left(\sum_{\ell=0}^N \gamma(N - \ell) E[T_\ell] \right) E[S_{TT}],$$

which, together with Eqs. (16) and (17), gives us that

$$\frac{1}{\lambda + \gamma(N)} = \frac{\pi_N}{\alpha \lambda R_{0N}} (1 - \lambda(E[S_{ET}] - E[S_{TT}])) - \frac{E[S_{TT}]}{\alpha}.$$

This can be rewritten to

$$\alpha \lambda R_{0N} = (\pi_N (1 - \lambda(E[S_{ET}] - E[S_{TT}])) - \lambda R_{0N} E[S_{TT}]) (\lambda + \gamma(N)).$$

Insert $\pi_N = \sigma_N + \alpha \tau_N$ (see Eq (14)) to find

$$\alpha = \frac{(\sigma_N (1 - \lambda(E[S_{ET}] - E[S_{TT}])) - \lambda R_{0N} E[S_{TT}]) (\lambda + \gamma(N))}{\lambda R_{0N} - \tau_N (1 - \lambda(E[S_{ET}] - E[S_{TT}])) (\lambda + \gamma(N))}.$$

So now we have the expressions for σ , τ , and α , and we can use Eq. (14) to determine steady state probabilities (π) of N_{TT} (the number of TTs at the ASC). ■

Appendix D. Modified AMVA algorithm for multiple ASCs

For now consider the system in Fig. 9a. Let the subnetwork contain J called internal stations, where TTs are handled one by one with a node dependent service time S_{Ij} on one of the c_j servers, $j = 1, \dots, J$. For this case, $J = 3$ corresponding to the Travel 1, GC, and Travel 2 node in Fig. 5b. ETs arrive at ASC k with rate λ_k ; the service times are $S_{ET,k}$ and $S_{TT,k}$, for $k = 1, \dots, K$. We denote the visit ratio of a TT to internal station j by V_j where the visit ratio is relative to one visit to any ASC. When a TT visits an ASC, the probability that ASC(k) is visited is p_k . Let $P(L_{Qj}(n) = m)$, $m = 0, \dots, n$ and $j = 1, \dots, J$ be the (approximate) queue length distribution at internal station j and $W_k(n)$ the waiting time of a TT at ASC(k) when n TTs are present.

The crucial step in the modified AMVA algorithm is to estimate the throughput time of a TT at a station. In the modified AMVA, we keep track of the marginal probability distribution of the number of TTs at the internal stations (as in the normal AMVA), and we use the exact analysis results for the expected throughput time at an ASC in isolation from Section 4. In our experiments, we observe convergence for Step 4 in the algorithm.

Modified AMVA algorithm

As initialization, let $P(L_{Qj}(0) = 0) = 1$, for $j = 1, \dots, J$ and $E[W_k(0)] = 0$, for $k = 1, \dots, K$ and set $n = 1$.

1. For $j = 1, \dots, J$, compute

$$E[W_{Ij}] = P(L_{Qj}(n-1) \geq c_j) E[S_{Ij}^r] + E[(L_{Qj}(n-1) - c_j)^+] \frac{E[S_{Ij}]}{c_j} + E[S_{Ij}]$$

where

$$E[S_{Ij}^r] = \frac{(c_j - 1)E^2[S_{Ij}] + E[S_{Ij}^2]}{c_j(c_j + 1)E[S_{Ij}]};$$

2. Set $W_{IJ} = \sum_{j=1}^J V_j E[W_{Ij}]$ and $E[W_k(n)] = E[W_k(n-1)]$, for $k = 1, \dots, K$;

3. Repeat

(a) Set $E[W_k^I] = E[W_k(n)]$, for $k = 1, \dots, K$;

(b) For $k = 1, \dots, K$, compute

$$TH(k, n) = \frac{np_k}{W_{IJ} + \sum_{\ell=1, \ell \neq k}^K p_\ell E[W_\ell(n)]};$$

(c) For $k = 1, \dots, K$, compute $E[W_k(n)]$ by the algorithm of the previous section with return rates $TH(k, m)$, $m = 1, \dots, n$;

Until $E[W_k(n)]$ is close to $E[W_k^I]$, for all $k = i, \dots, K$;

4. Compute $TH(n) = \frac{n}{W_{IJ} + \sum_{k=1}^K p_k E[W_k(n)]}$;

5. Set

$$P(L_{Q_j}(n) = m + 1) = \begin{cases} P(L_{Q_j}(n-1) = m) V_j TH(n) \frac{E[S_{I_j}]}{m+1} & \text{for } m = 0, \dots, c_j - 1 \\ P(L_{Q_j}(n-1) = m) V_j TH(n) \frac{c_j E[S_{I_j}^r] + m E[S_{I_j}]}{(m+1)c_j} & \text{for } m = c_j, \dots, n-1, \end{cases}$$

and

$$P(L_{Q_j}(n) = 0) = 1 - P(L_{Q_j}(n) \geq 0);$$

6. Increase n ;

7. Repeat the previous steps until $n > N$.

Lemma 17. Under the following conditions, Step 3 in the modified AMVA algorithm converges: (1) if each stack has $1 + N$ cranes and exactly one crane is used to handle the external truck (all cranes can pass each other without blocking), (2) if we have a system of N TTs and N ASCs, each stack has two ASCs with one dedicated to TT and another dedicated to ET, and each TT is assigned to only one ASC.

Proof. To prove the convergence for these special cases is easy, since $E[W_{\ell}(n)] = S_{TT,\ell}$ and so $TH(k, n) = \frac{np_k}{W_{IJ} + \sum_{l=1, l \neq k}^K p_l S}$ which does not change after the first iteration. ■

Appendix E. Model validation for coupled system without train station

To validate the network approximations when the TTs are always busy (i.e., a train container is always available for unloading), we design a set of experiments. They are: (1) Network with one ASC and exponential return times for the TTs to the ASC (no additional queues); (2) Network with one ASC and deterministic return times for the TTs to the ASC (no additional queues); (3) Network with one ASC, one GC single server queue with general service times (for container unloading at the trainside), and exponential return times for the TTs to the ASC; (4) Network with four and five ASCs, one GC single server queue with general service times (for container unloading at the trainside), and exponential return times for the TTs to the ASC; and (5) Network with eight and 12 ASCs and general service times, one GC with uniform travel times, and exponential return times for the TTs to the ASC. Tables E.7 and E.11 show the results for these five cases. The performance measures that are of interest to us include utilization of the ASC resource (U_{ASC}), expected waiting time for the ETs before service ($E[W_{ET}]$), expected waiting time for the TTs before service ($E[W_{TT}]$), overall (round trip) expected throughput time for the TT ($E[T_{TT,O}]$), throughput of the TT (TH_{TT}), utilization of the GC (U_{GC}), and expected queue length at the ASC and the GC resource (denoted by $E[L_{ASC}]$ and $E[L_{GC}]$ respectively). We additionally denote the source of the measure, which is either analytical, A or simulation, S . We simulate the closed queuing network with ET arrivals at the ASC using a discrete-event simulation software, Arena. We run the simulation for 25 days and 15 replications. The 95% confidence intervals are within 5% of the averages.

Table E.7 presents the results based on our exact analysis. So it is not surprising to note that the validation results are quite good (the absolute percentage error for a measure x , $|\frac{A(x) - S(x)}{S(x)}|$, is less than 0.5%). When the exponential return times are modified to deterministic return times in Table E.8, the absolute percentage errors for the performance measures increase, in particular the expected waiting time errors for the ETs and the TTs increase up to 30% and 20%, respectively. The absolute percentage error for the expected queue length at the ASC also increases to 45%. Likewise, when we introduce one GC and one ASC in the system, the absolute percentage errors for the expected waiting time for the ETs and the TTs increase up to 40% and 8%, respectively (Table E.9). The absolute percentage error for the expected queue length at the ASC is about 30%. However, the absolute percentage errors in the utilization measures are quite low (about 0.07% in Tables E.7 and E.9 and about 2% in Table E.8). Note that Tables E.7–E.9 represent small-scale problem instances with one ASC. In practical systems, we have interactions with four to 12 ASCs. We test the model performance for network with large number of ASCs (4–5 and 8–12) in Tables E.10 and E.11, respectively. In such systems, we observe that the percentage errors for the performance measures especially for the expected queue length errors and therefore the external waiting times for the ETs and TTs reduce to about 15%. This trend is observed because in large scale networks the arrival streams to the stations are random and not quite deterministic (which was the situation for one ASC and one GC network). We develop the model for large-scale ASC systems with the train station and test the performance in Sections 6 and 7, respectively.

Appendix F. Decoupled system performance analysis with train arrivals

We first discuss the integrated open queuing network model of the decoupled system as illustrated in Fig. 14. For the decoupled system, we list down the key modeling assumptions: 1. ETs arrive at the terminal according to a Poisson Process with rate λ (we use λ_k for ET arrivals at ASC k); 2. A train container has a non-preemptive service priority over an ET container; 3. The GC begins unload operations as soon as train containers are available. However, ASCs begin their operations when either a train container or an ET is present in the ASC buffer lane; 4. The train containers arrive in a batch renewal process with rate (λ_T). The interarrival times of the trains are deterministic; 5. The processing times at the GC and ASCs follow general distributions. The train containers queue at the GC station for unloading. Then the containers queue at the decoupled LV multi-server station for transport. Finally,

Table E.7

Single ASC case with 4 TTs and with an average 15 and 25 min Exponential delay.

Delay	λ (trucks/h)	S/A	U_{ASC}	$E[W_{ET}]$ (s)	$E[W_{TT}]$ at ASC (s)	TH_{TT} (trips per h)	$E[T_{TT,O}]$ (s)	$E[L_{ASC}]$
15	4.8	S	84.0%	880.3	300.2	11.99	1200.6	1.333
		A	84.0%	885.1	299.9	12.00	1199.9	1.340
	5.5	S	87.0%	1077.7	305.9	11.96	1204.0	1.772
		A	87.0%	1081.7	306.4	11.94	1206.4	1.785
	6.3	S	90.9%	1528.7	315.2	11.86	1213.8	2.808
		A	90.8%	1533.4	315.0	11.85	1215.0	2.819
25	4.8	S	64.9%	407.4	254.9	8.20	1756.2	0.473
		A	65.0%	409.8	255.1	8.20	1755.1	0.478
	5.5	S	68.0%	444.9	259.8	8.17	1761.5	0.581
		A	68.2%	446.4	260.0	8.18	1760.0	0.585
	6.3	S	72.3%	505.7	265.9	8.15	1766.2	0.766
		A	72.3%	507.4	266.4	8.15	1766.4	0.770

Table E.8

Single ASC case with 4 TTs and with an average 15 and 25 min Deterministic Delay.

Delay	λ (trucks/h)	S/A	U_{ASC}	$E[W_{ET}]$ (s)	$E[W_{TT}]$ at ASC (s)	TH_{TT} (trips per h)	$E[T_{TT,O}]$ (s)	$E[L_{ASC}]$
15	4.8	S	87.0%	673.0	243.9	12.60	1143.9	0.891
		A	84.0%	885.1	299.9	12.00	1199.9	1.340
	5.5	S	89.1%	841.9	252.9	12.50	1152.9	1.268
		A	87.0%	1081.7	306.4	11.94	1206.4	1.785
	6.3	S	93.5%	1251.8	265.0	12.37	1165.0	2.197
		A	90.8%	1533.4	315.0	11.85	1215.0	2.819
25	4.8	S	65.9%	342.4	217.5	8.39	1717.5	0.304
		A	65.0%	409.8	255.1	8.20	1755.1	0.478
	5.5	S	68.9%	370.1	222.9	8.37	1722.9	0.389
		A	68.2%	446.4	260.0	8.18	1760.0	0.585
	6.3	S	73.3%	417.8	230.1	8.33	1730.1	0.537
		A	72.3%	507.4	266.4	8.15	1766.4	0.770

Table E.9

Single ASC case with one GC (Uniform - (3,5) min) and Exponential delay of 7.5 min before and after GC usage.

N	λ (trucks/h)	S/A	U_{ASC}	$E[W_{ET}]$ (s)	$E[W_{TT}]$ (s)	$E[T_{TT,O}]$ (s)	TH_{TT} (trips/h)	U_{GC}	$E[L_{ASC}]$	$E[L_{GC}]$
4	4.8	S	72.7%	444.9	258.1	1483.2	9.7	64.7%	0.565	0.225
		A	72.6%	525.7	271.9	1503.9	9.6	63.8%	0.705	0.308
	5.5	S	75.9%	495.6	264.3	1487.2	9.7	64.5%	0.709	0.226
		A	75.7%	589.3	277.3	1508.9	9.5	63.6%	0.878	0.310
	6.3	S	80.2%	589.6	271.5	1494.3	9.6	64.2%	0.971	0.230
		A	79.8%	702.7	284.4	1515.5	9.5	63.3%	1.193	0.305
5	4.8	S	82.0%	664.9	283.6	1556.9	11.6	77.0%	0.985	0.436
		A	82.0%	860.1	306.4	1591.0	11.3	75.4%	1.304	0.544
	5.5	S	85.0%	736.7	289.6	1561.5	11.5	76.8%	1.202	0.430
		A	85.1%	1035.3	312.7	1596.6	11.3	75.2%	1.712	0.539
	6.3	S	88.9%	996.6	299.7	1571.9	11.5	76.4%	1.814	0.420
		A	89.1%	1416.8	321.2	1604.0	11.2	74.8%	2.610	0.533

the containers are unloaded and stored at the ASC single-server station buffer. The ETs arrive at an ASC (k), according to a Poisson Process with rate λ_k .

We first give a necessary and sufficient condition for stability based on the number of decoupled trucks can be obtained by modeling the system with a $GI/G/N_{LV}$ multi-server queue, where N_{LV} denotes the number of decoupled truck. Here the coordination delays are absent as the vehicle can unload and load the container from the ground by its own.

We provide the model notations in Table F.12.

Lemma 18. A necessary and sufficient condition for subnetwork stability is given by $N_{LV} > (N_{CT}\lambda_T)(2T_{LV} + 2T_{UL})$.

Table E.10

Four and five ASC case with one GC (Uniform - (3,5) min), service time at ASC Uniform (120,240 s), and Exponential delay (with mean 7.5 min) before the GC service and another Exponential delay (with mean 7.5 min) after the GC service; all time measures are reported in minutes.

N_{ASC}	N	λ	S/A	U_{ASC}	$E[W_{ET}]$	$E[W_{TT}]$	$E[T_{TT,O}]$	TH_{TT}	U_{GC}	$E[L_{ASC}]$	$E[L_{GC}]$
4	4	48.0	S	72.2%	7.52	4.22	24.74	9.70	64.7%	1.101	0.246
			A	72.0%	7.45	4.19	24.64	9.74	64.9%	0.924	0.235
		60.0	S	87.1%	14.73	4.48	24.98	9.61	64.0%	3.171	0.240
			A	87.0%	14.40	4.45	24.86	9.66	64.4%	2.818	0.222
		66.7	S	95.3%	38.58	4.63	25.12	9.55	63.7%	10.143	0.237
			A	95.2%	38.05	4.61	24.99	9.61	64.0%	10.095	0.225
4	5	48.0	S	74.7%	8.28	4.30	25.74	11.66	77.7%	1.309	0.473
			A	74.7%	8.15	4.26	25.60	11.72	78.1%	1.093	0.462
		60.0	S	89.5%	18.34	4.58	25.97	11.55	77.0%	4.139	0.461
			A	89.5%	17.87	4.54	25.87	11.60	77.3%	3.872	0.447
		66.7	S	97.8%	82.67	4.73	26.10	11.49	76.6%	22.461	0.455
			A	97.7%	79.64	4.69	25.98	11.54	77.0%	20.180	0.442
5	5	60.0	S	72.7%	7.69	4.24	25.68	11.68	68.5%	1.179	0.569
			A	71.7%	7.37	4.19	25.60	11.72	78.1%	0.919	0.465
		75.0	S	87.6%	15.41	4.51	25.91	11.58	68.0%	3.394	0.556
			A	86.5%	14.19	4.46	25.77	11.64	77.6%	2.877	0.448

Table E.11

Eight and twelve ASC case with one GC (Uniform - (3,5) min), service time at ASC Uniform (120,240 s), and Exponential delay (with mean 7.5 min) before the GC service and another Exponential delay (with mean 7.5 min) after the GC service; all time measures are reported in minutes.

N_{ASC}	N	λ	S/A	U_{ASC}	$E[W_{ET}]$	$E[W_{TT}]$	$E[T_{TT,O}]$	TH_{TT}	U_{GC}	$E[L_{ASC}]$	$E[L_{GC}]$
8	8	96.0	S	73.3%	7.9	4.3	22.5	21.3	71.2%	1.030	0.447
			A	73.3%	7.9	4.3	22.6	21.2	70.7%	1.032	0.557
		120.0	S	88.2%	16.2	4.5	22.8	21.1	70.3%	3.376	0.431
			A	88.2%	16.2	4.5	22.9	21.0	70.0%	3.373	0.541
		133.3	S	96.4%	50.4	4.7	22.9	21.0	69.9%	13.235	0.423
			A	96.4%	50.7	4.7	23.0	20.9	69.6%	13.331	0.532
12	12	145.2	S	74.2%	8.2	4.3	21.7	33.2	73.7%	1.092	0.591
			A	73.8%	8.0	4.3	21.8	33.0	73.4%	1.069	0.723
		181.8	S	88.7%	17.0	4.6	22.0	32.8	72.9%	3.572	0.570
			A	88.6%	17.0	4.6	22.0	32.7	72.6%	3.564	0.700
		200.0	S	96.8%	57.6	4.7	22.1	32.6	72.5%	16.002	0.560
			A	96.9%	58.8	4.7	22.2	32.5	72.2%	15.565	0.688

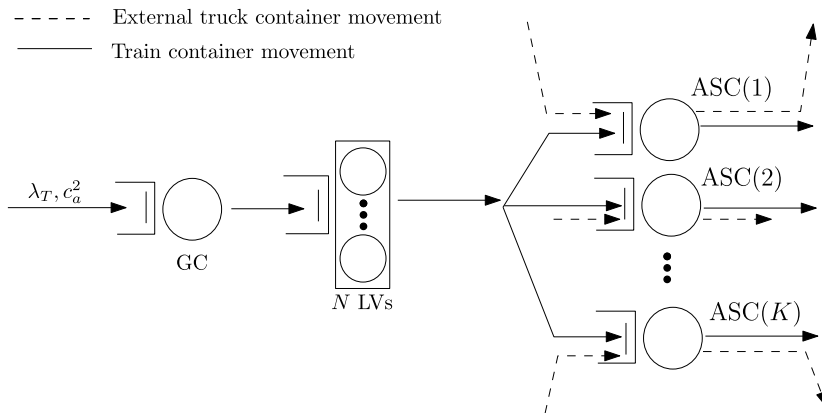
**Fig. 14.** Integrated model for the decoupled system with multiple ASCs.

Table F.12
Notations for the decoupled system model.

Term	Description
N_{LV}	Number of LVs in the system
T_{LV}	Travel time of an LV from the GC to an ASC or vice versa
T_{UL}	LV pickup time and set down time of the container on the ground
N_{CT}	Number of containers on the train
T_{CP}	Time taken by the GC to dropoff a container at the buffer location
T_{CS}	Travel time between two consecutive container loads on the train
$E[W_{GC}]$	Expected waiting time of a container before the GC picks up the container
$TT_{GC,n}$	Throughput time of the n th container by the GC
N_{ET}	Number of External trucks at an ASC
T_{REM}	Remaining ASC service time at the moment of an LV arrival

Proof. From the stability condition for a $GI/G/N_{LV}$ queue, $N_{CT}\lambda_T < \frac{N_{LV}}{(2T_{LV}+2T_{UL})}$, Lemma 18 follows. ■

Analysis of the train station

In this case, we do not use the analysis of the train station in Section 6. An essential assumption was exponential return times of LVs to the GC. Here the return time is deterministic, namely the sum of the two deterministic travel times from the ASC to the GC and back. The train interarrival time is also assumed to be deterministic. The stability of the system now also requires that a train is fully unloaded before another train arrives, i.e., all LVs are free at the arrival instant of a train. For the analysis of an ASC, we use a non-homogeneous Markov chain and compute the transient behavior.

In the decoupled system, the GC unloads each container in a sequential order. The GC movement is sequential for the decoupled resources, i.e., the containers are picked up from the first container location on the train and the GC moves sequentially towards the last container that needs unloading. The speed of the GC determines the unloading of the train. Throughout this section, we assume both deterministic handling times at the GC and deterministic travel times. We can easily see that the total unloading time of a train is given by $N_{CT}T_{CP} + (N_{CT} - 1)T_{CS}$, where the crane is assumed to be in the first position. T_{CP} and T_{CS} denote the container pickup time from the train by the GC and the travel time between two consecutive container slots, respectively. The total time for the crane to return to the first position equals the unloading time of the train plus the return time to the first position, and is given by $N_{CT}T_{CP} + 2(N_{CT} - 1)T_{CS}$. The throughput time at the GC of the n th container on a train ($n \leq N_{CT}$) is $TT_{GC,n} = (n - 1)(T_{CP} + T_{CS}) + T_{CP}$, which gives an average waiting time of

$$E[W_{GC}] = \frac{1}{N_{CT}} \sum_{n=1}^{N_{CT}} TT_{GC,n} = \frac{1}{2}(N_{CT} - 1)(T_{CP} + T_{CS}) + T_{CP}.$$

Analysis of the lifting vehicle station

After the container is unloaded from the train it might have to wait for an LV, depending on the return travel time of an LV between the GC and the ASC. Denote the travel time of an LV from an ASC to GC or vice-versa by T_{LV} . Since all LVs are free when the train arrives, the waiting time for the first N containers is T_{LV} and the total time until delivery at the ASC equals

$$T_{LV} + T_{UL} + T_{LV} = 2T_{LV} + T_{UL},$$

where T_{UL} is the LV pickup time from the ground. Let us use an index n for the LVs according to the sequence of container pickup from the GC buffer, where $n \in \{1, \dots, N\}$. The first LV picks up a container at time $T_{CP} + T_{LV}$, which corresponds to the time it takes the GC to dropoff a container at the buffer location, and the time for an LV to travel from the ASC to the GC buffer to pick up the container. This LV is free at time $2T_{LV} + 2T_{UL} + T_{CP}$. The $N + 1$ -th container is dropped off at the GC buffer location at time $T_{CP} + N(T_{CP} + T_{CS})$. If this event occurs after the first LV is free, then the waiting time of this container is just T_{LV} . Otherwise, it is $2T_{LV} + 2T_{UL} + T_{CP} - (T_{CP} + N(T_{CP} + T_{CS})) + T_{LV}$. Equivalently, we could say that the waiting time is $[2T_{LV} + 2T_{UL} - N(T_{CP} + T_{CS})]^+ + T_{LV}$. A similar argument applies for the n th LV case and so the $N + 1$ -th container has the same waiting time for $n \in \{1, \dots, N\}$. Continuing this argumentation gives the waiting time of the $n + kN$ -th container as

$$k[2T_{LV} + 2T_{UL} - N(T_{CP} + T_{CS})]^+ + T_{LV}.$$

We can also estimate the average container waiting time for an LV.

Analysis of the ASCs

We assume that the distribution of the handling times at an ASC is the same for LVs and ETs. To analyze the waiting times at an ASC, we introduce a discrete time three dimensional non-homogeneous Markov chain for the number of LVs, ETs at the ASC, and the remaining service time (with state space $\{(N_{LV}, N_{ET}, T_{REM})\}$) and iterate over a period containing several train arrivals. We assume that arrivals of LVs at ASC (k) can only occur at time points determined by the departure process from the GC with probability p_k . On the other hand, ETs can arrive at every time point with a fixed probability since in real time we have a Poisson arrival process. Note that we ignore that more than one external truck can arrive in the discrete time interval. The probability that an ET arrives at a certain time point is $\lambda_k \Delta$, where Δ is the time discretization. If a service starts, T_{REM} is set to the total service

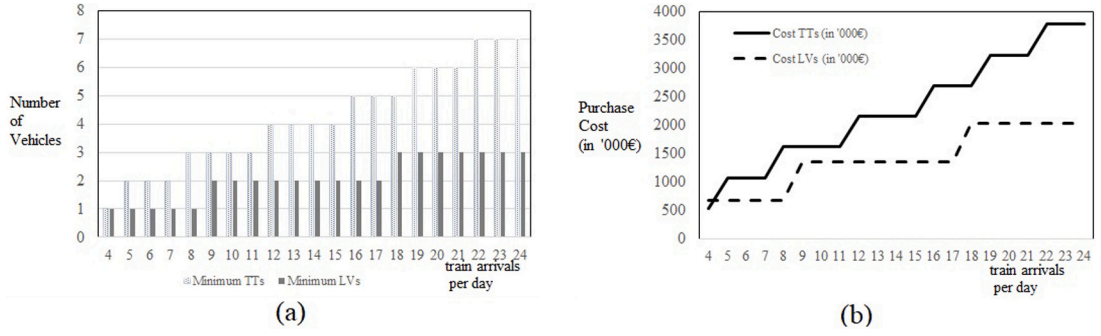


Fig. 15. (a) Minimum number of vehicles for 8 train arrivals per day and 130 trucks per hour (b) Cost of vehicle purchase.

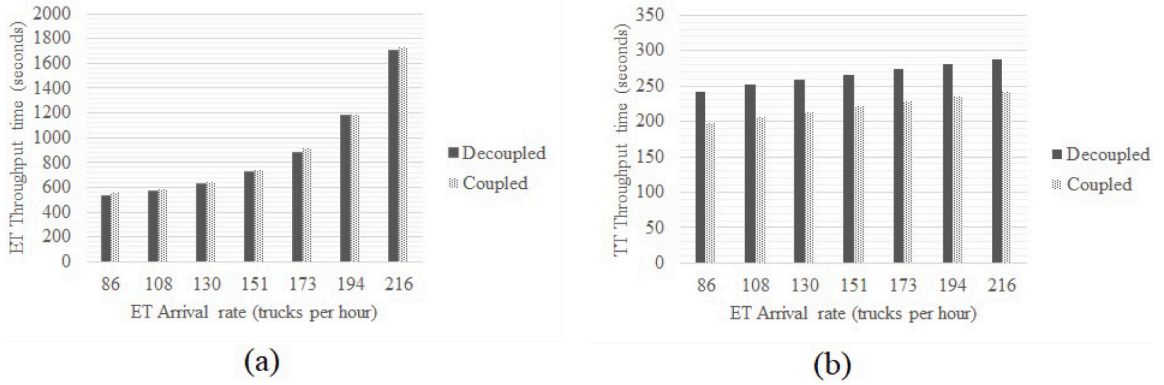


Fig. 16. (a) ET throughput times (b) TT throughput times.

time. Every time step T_{REM} is decreased by one until it reaches zero. Then a new service starts when $N_{LV} + N_{ET} > 0$, taking into account the priorities.

Once we have the expected number of LVs and ETs, we can use Little's Law to find the expected waiting time, both for LVs and ETs.

Appendix G. Cost-time trade-offs in vehicle selection

First, we identify the minimum number of vehicles required in the coupled and decoupled system with varying train arrival rates to make the system stable. We translate the minimum number of vehicles required to equivalent cost. We set the purchase cost per (coupled) TT to €540 K and the cost per (decoupled) LV ranges to €680 K (source: <https://tba.group/en>). We fix the ET arrival rate to 130 trucks per hour. From Fig. 15, we find that when the train arrival rates are quite low, TTs (coupled) vehicles offer a lower cost option; however, when the train arrival rates are high, LVs (decoupled) vehicles offer a lower cost option (as fewer vehicles are required), see the graphs corresponding to the minimum number of vehicles and the corresponding costs. Our results also suggest that, for a given number of vehicles, LVs offer significantly lower ET throughput times at an expense of slightly higher TT throughput time. Hence, decoupled vehicles seem to be a good choice from both cost and time perspective (especially at high train arrival rates).

As an example, consider a scenario for a decoupled system with 1 LV (86% utilization), 8 trains arriving per day and 130 trucks arrive per hour, the average ET and TT throughput time at the ASC are 564 s and 202 s, respectively. A corresponding scenario for a coupled system needs a minimum of 2 TTs (leading to 82% utilization), with 8 trains arriving per day and 130 trucks arriving per hour. The resulting average ET and TT throughput time at the ASC are 582 s and 185 s, respectively. If the terminal strives for target average throughput times of 600 s and 210 s, for ETs and TTs, respectively, then the terminal can select one LV with a purchase cost of €680 K instead of two TTs with a purchase cost of €1080 K.

Appendix H. Comparison of system performance with time-varying arrivals

For the train schedule, there is some level of uncertainty in the last-mile and hence we model the train arrivals with a renewal process with known mean and CV of inter-arrival times (albeit low value, i.e. small variation only). The arrival of trucks could strongly vary by day along the week and by hour of the day in most sea terminal containers. Since our interest is to model the

interaction between trucks and train containers at the ASC, we use a time-homogeneous Poisson process (constant rate) to model the ET arrivals. In this section, we use simulation to confirm that the insights from the model with constant arrival rate also hold true if we use a Markov-Modulated Poisson process for the ET arrivals, with two levels of arrival rates, both with exponential periods of duration. We also consider six additional gates for the ETs to enter the terminal with time-varying arrival rates. The average demand rate during peak times is two times the demand during normal times and they last for one hour windows in a day. We choose this configuration so that existing container terminal equipment can be used without creating network stability issues.

For a configuration with 10 internal transport vehicles, 14 ASCs, 1 GC, and 8 trains arriving per day, We find that the insights from the model are similar to those obtained with the stationary Poisson arrival process (see Fig. 16). In particular, We find that the ET throughput time at the ASC is slightly longer for the coupled system in comparison to the decoupled system (up to 4%). However, the TT throughput time improves significantly with the coupled system compared with the decoupled system (up to 22%).

References

- Ahern, Z., Paz, A., Corry, P., 2022. Approximate multi-objective optimization for integrated bus route design and service frequency setting. *Transp. Res. B* 155, 1–25.
- Ambrosino, D., Siri, S., 2015. Comparison of solution approaches for the train load planning problem in seaport terminals. *Transp. Res. E* 79, 65–82.
- Azab, A., Morita, H., 2022. The block relocation problem with appointment scheduling. *European J. Oper. Res.* 297 (2), 680–694.
- Borndörfer, R., Klug, T., Lamorgese, L., Mannino, C., Reuther, M., Schlechte, T., 2018. *Handbook of Optimization in the Railway Industry*, Vol. 268. Springer.
- Boysen, N., Flidner, M., 2010. Determining crane areas in intermodal transshipment yards: The yard partition problem. *European J. Oper. Res.* 204 (2), 336–342.
- Boysen, N., Flidner, M., Jaehn, F., Pesch, E., 2012a. Shunting yard operations: Theoretical aspects and applications. *European J. Oper. Res.* 220 (1), 1–14.
- Boysen, N., Jaehn, F., Pesch, E., 2011. Scheduling freight trains in rail-rail transshipment yards. *Transp. Sci.* 45 (2), 199–211.
- Boysen, N., Jaehn, F., Pesch, E., 2012b. New bounds and algorithms for the transshipment yard scheduling problem. *J. Sched.* 15 (4), 499–511.
- Bruck, B.P., Cordeau, J.-F., Frejinger, E., 2021. Integrated inbound train split and load planning in an intermodal railway terminal. *Transp. Res. B* 145, 270–289.
- Buzacott, J., Shanthikumar, J., 1993. *Stochastic Models of Manufacturing Systems*. Prentice Hall.
- Bychkov, I., Kazakov, A., Lempert, A., Zharkov, M., 2021. Modeling of railway stations based on queueing networks. *Appl. Sci.* 11 (5), 2425.
- Carlo, H., Vis, I., Roodbergen, K., 2013. Storage yard operations in container terminals: Literature overview, trends, and research directions. *European J. Oper. Res.*
- Chen, H., Lam, J.S.L., Liu, N., 2018. Strategic investment in enhancing port–hinterland container transportation network resilience: A network game theory approach. *Transp. Res. B* 111, 83–112.
- Chen, S., Wang, H., Meng, Q., 2021. Autonomous truck scheduling for container transshipment between two seaport terminals considering platooning and speed optimization. *Transp. Res. B* 154, 289–315.
- Chen, G., Yang, Z., 2010. Optimizing time windows for managing export container arrivals at Chinese container terminals. *Marit. Econ. Logist.* 12 (1), 111–126.
- Corry, P., Bierwirth, C., 2019. The berth allocation problem with channel restrictions. *Transp. Sci.* 53 (3), 708–727.
- Dhingra, V., Kumawat, G.L., Roy, D., De Koster, R., 2018. Solving semi-open queueing networks with time-varying arrivals: An application in container terminal landside operations. *European J. Oper. Res.* 267 (3), 855–876.
- Dhingra, V., Roy, D., De Koster, R., 2017. A cooperative quay crane-based stochastic model to estimate vessel handling time. *Flex. Serv. Manuf. J.* 29 (1), 97–124.
- Dorda, M., Teichmann, D., 2013. Modelling of freight trains classification using queueing system subject to breakdowns. *Math. Probl. Eng.* 2013.
- Duru, O., Galvao, C.B., Mileski, J., Robles, L.T., Gharehgozli, A., 2020. Developing a comprehensive approach to port performance assessment. *Asian J. Ship. Logist.* 36 (4), 169–180.
- Europe Container Terminals, 2015. Euromax terminal rotterdam - safe and secure. Accessed on 12.06.2015.
- Frisch, S., Hungerländer, P., Jellen, A., Primas, B., Steininger, S., Weinberger, D., 2021. Solving a real-world locomotive scheduling problem with maintenance constraints. *Transp. Res. B* 150, 386–409.
- Gharehgozli, A.H., De Koster, R., Jansen, R., 2017. Collaborative solutions for inter terminal transport. *Int. J. Prod. Res.* 55 (21), 6527–6546.
- Gharehgozli, A., Roy, D., De Koster, R., 2015. Sea container terminals: New technologies and or models. *Marit. Econ. Logist.* 30 (1).
- Gharehgozli, A., Zaepour, N., De Koster, R., 2020. Container terminal layout design: transition and future. *Marit. Econ. Logist.* 22 (4), 610–639.
- Giuliano, G., O'Brien, T., 2007. Reducing port-related truck emissions: The terminal gate appointment system at the ports of los angeles and long beach. *Transp. Res. D* 460–473.
- Gorman, M.F., Clarke, J.-P., Gharehgozli, A.H., Hewitt, M., De Koster, R., Roy, D., 2014. State of the practice: A review of the application of or/ms in freight transportation. *INFORMS J. App. Anal.* 44 (6), 535–554.
- Guan, C., Liu, R., 2009. Container terminal gate appointment system optimization. *Marit. Econ. Logist.* 11 (4), 378–398.
- Hoshino, S., Ota, J., Shinozaki, A., Hashimoto, H., 2007. Optimal design methodology for an agv transportation system by using the queueing network theory. In: *Distributed Autonomous Robotic Systems*, vol. 6. Springer, pp. 411–420.
- Hu, Q., Wiegman, B., Corman, F., Lodewijks, G., 2019. Integration of inter-terminal transport and hinterland rail transport. *Flex. Serv. Manuf. J.* 31 (3), 807–831.
- Huisman, T., Boucherie, R.J., 2001. Running times on railway sections with heterogeneous train traffic. *Transp. Res. B* 35 (3), 271–292.
- Huisman, T., Boucherie, R.J., Van Dijk, N.M., 2002. A solvable queueing network model for railway networks and its validation and applications for the netherlands. *European J. Oper. Res.* 142 (1), 30–51.
- Jaehn, F., 2013. Positioning of loading units in a transshipment yard storage area. *Or Spectrum* 35 (2), 399–416.
- Jia, J., Heragu, S.S., 2009. Solving semi-open queueing networks. *Oper. Res.* 57 (2), 391–401.
- Kang, S., Medina, J.C., Ouyang, Y., 2008. Optimal operations of transportation fleet for unloading activities at container ports. *Transp. Res. B* 42 (10), 970–984.
- Kizilay, D., Hentenryck, P.Van, Eliyi, D.T., 2020. Constraint programming models for integrated container terminal operations. *European J. Oper. Res.* 286 (3), 945–962.
- Krishnamurthy, A., Suri, R., Vernon, M., 2004. Analysis of a fork/join synchronization station with inputs from coxian servers in a closed queueing network. *Ann. Oper. Res.* 125 (1), 69–94.
- Lai, Y.-C., Barkan, C.P., Önal, H., 2008. Optimizing the aerodynamic efficiency of intermodal freight trains. *Transp. Res. E* 44 (5), 820–834.
- Leachman, R.C., Julia, P., 2012. Estimating flow times for containerized imports from asia to the united states through the western rail network. *Transp. Res. E* 48 (1), 296–309.
- Lee, B.K., Lee, L.H., Chew, E.P., 2014. Analysis on container port capacity: A Markovian modeling approach. *OR Spectrum* 36 (2).
- Li, D., Dong, J.-X., Song, D.-P., Hicks, C., Singh, S.P., 2020a. Optimal contract design for the exchange of tradable truck permits at multiterminal ports. *Int. J. Prod. Econ.* 230, 107815.
- Li, K., Gharehgozli, A., Ahuja, M.V., Lee, J.-Y., 2020b. Blockchain in maritime supply chain: A synthesis analysis of benefits, challenges and limitations. *J. Supply Chain Oper. Manag.* 18 (2), 257–273.

- Li, K., Lee, J.-Y., Gharehgozli, A., 2021. Blockchain in food supply chains: a literature review and synthesis analysis of platforms, benefits and challenges. *Int. J. Prod. Res.* 1–20.
- Liu, M., Lee, C.-Y., Zhang, Z., Chu, C., 2016. Bi-objective optimization for the container terminal integrated planning. *Transp. Res. B* 93, 720–749, Maritime Logistics.
- Liu, B., Li, Z.-C., Sheng, D., Wang, Y., 2021. Integrated planning of berth allocation and vessel sequencing in a seaport with one-way navigation channel. *Transp. Res. B* 143, 23–47.
- Mantovani, S., Morganti, G., Umang, N., Crainic, T.G., Frejinger, E., Larsen, E., 2018. The load planning problem for double-stack intermodal trains. *European J. Oper. Res.* 267 (1), 107–119.
- Mar-Ortiz, J., Castillo-García, N., Gracia, M.D., 2020. A decision support system for a capacity management problem at a container terminal. *Int. J. Prod. Econ.* 222, 107502.
- Meisel, F., Bierwirth, C., 2013. A framework for integrated berth allocation and crane operations planning in seaport container terminals. *Transp. Sci.* 47 (2), 131–147.
- Mishra, N., Roy, D., van Ommen, J.-K., 2017. A stochastic model for interterminal container transportation. *Transp. Sci.* 51 (1), 67–87.
- Ng, M., Talley, W.K., 2020. Rail intermodal management at marine container terminals: Loading double stack trains. *Transp. Res. C* 112, 252–259.
- Özkan, E., Ward, A.R., 2019. On the control of fork-join networks. *Math. Oper. Res.* 44 (2), 532–564.
- Phan, M.-H., Kim, K.H., 2016. Collaborative truck scheduling and appointments for trucking companies and container terminals. *Transp. Res. B* 86, 37–50.
- Roy, D., 2016. Semi-open queueing networks: a review of stochastic models, solution methods and new research areas. *Int. J. Prod. Res.* 54 (6), 1735–1752.
- Roy, D., De Koster, R., 2018. Stochastic modeling of unloading and loading operations at a container terminal using automated lifting vehicles. *European J. Oper. Res.* 266 (3), 895–910.
- Roy, D., De Koster, R., Bekker, R., 2020. Modeling and design of container terminal operations. *Oper. Res.* 68 (3), 686–715.
- Roy, D., Gupta, A., De Koster, R., 2016. A non-linear traffic flow-based queueing model to estimate container terminal throughput with AGVs. *Int. J. Prod. Res.* 54 (2), 472–493.
- Ruf, M., Cordeau, J.-F., 2021. Adaptive large neighborhood search for integrated planning in railroad classification yards. *Transp. Res. B* 150, 26–51.
- Rupp, J., Boysen, N., Briskorn, D., 2021. Optimizing consolidation processes in hubs: The hub-arrival-departure problem. *European J. Oper. Res.*
- Saini, S., Roy, D., De Koster, R., 2017. A stochastic model for the throughput analysis of passing dual yard cranes. *Comput. Oper. Res.* 87 (C), 40–51.
- Schulz, A., Flidner, M., Friedrich, B., Pfeiffer, C., 2021. Levelling crane workload in multi-yard rail-road container terminals. *European J. Oper. Res.* 293 (3), 941–954.
- Torkjazi, M., Huynh, N., Shiri, S., 2018. Truck appointment systems considering impact to drayage truck tours. *Transp. Res. E* 116, 208–228.
- Turnquist, M.A., Daskin, M.S., 1982. Queueing models of classification and connection delay in railyards. *Transp. Sci.* 16 (2), 207–230.
- Upadhyay, A., 2020. Improving intermodal train operations in indian railways. *INFORMS J. Appl. Anal.* 50 (4), 213–224.
- Upadhyay, A., Gu, W., Bolia, N., 2017. Optimal loading of double-stack container trains. *Transp. Res. E* 107, 1–22.
- Vacca, I., Salani, M., Bierlaire, M., 2013. An exact algorithm for the integrated planning of berth allocation and quay crane assignment. *Transp. Sci.* 47 (2), 148–161.
- Wang, J., Liu, J., Wang, F., Yue, X., 2021a. Blockchain technology for port logistics capability: Exclusive or sharing. *Transp. Res. B* 149, 347–392.
- Wang, Y., Zhao, K., D'Ariano, A., Niu, R., Li, S., Luan, X., 2021b. Real-time integrated train rescheduling and rolling stock circulation planning for a metro line under disruptions. *Transp. Res. B* 152, 87–117.
- Wang, K., Zhen, L., Wang, S., Laporte, G., 2018. Column generation for the integrated berth allocation, quay crane assignment, and yard assignment problem. *Transp. Sci.* 52 (4), 812–834.
- Weik, N., Niebel, N., Nießen, N., 2016. Capacity analysis of railway lines in germany—a rigorous discussion of the queueing based approach. *J. Rail Transp. Plan. Manag.* 6 (2), 99–115.
- Wibowo, B., Fransoo, J., 2021. Joint-optimization of a truck appointment system to alleviate queueing problems in chemical plants. *Int. J. Prod. Res.* 59 (13), 3935–3950.
- Xie, Y., Song, D.-P., 2018. Optimal planning for container prestaging, discharging, and loading processes at seaport rail terminals with uncertainty. *Transp. Res. E* 119, 88–109.
- Yang, C.-S., 2019. Maritime shipping digitalization: Blockchain-based technology applications, future improvements, and intention to use. *Transp. Res. E* 131, 108–117.
- You, J., Miao, L., Zhang, C., Xue, Z., 2020. A generic model for the local container drayage problem using the emerging truck platooning operation mode. *Transp. Res. B* 133, 181–209.
- Zehndner, E., Feillet, D., 2014. Benefits of a truck appointment system on the service quality of inland transport modes at a multimodal container terminal. *European J. Oper. Res.* 235 (2), 461–469.
- Zeng, Y., Chaintreau, A., Towsley, D., Xia, C.H., 2018a. Throughput scalability analysis of fork-join queueing networks. *Oper. Res.* 66 (6), 1728–1743.
- Zeng, Y., Tan, J., Xia, C.H., 2018b. Fork and join queueing networks with heavy tails: Scaling dimension and throughput limit. *SIGMETRICS Perform. Eval. Rev.* 46 (1), 122–124.
- Zhang, Y., Peng, Q., Lu, G., Zhong, Q., Yan, X., Zhou, X., 2022. Integrated line planning and train timetabling through price-based cross-resolution feedback mechanism. *Transp. Res. B* 155, 240–277.
- Zhen, L., 2016. Modeling of yard congestion and optimization of yard template in container ports. *Transp. Res. B* 90, 83–104.