CIRRELT

CIRRELT-2024-12

Consolidation-based Transportation Planning: The Service Network Design Methodology

Teodor Gabriel Crainic

May 2024

Bureau de Montréal
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

Consolidation-based Transportation Planning: The Service Network Design Methodology

Teodor Gabriel Crainic*

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and School of Management, Université du Québec à Montréal

Abstract. Freight transportation supports trade exchanges generated by the needs and requirements of people, and, hence, of all public and private organizations making up the human society. Consolidation is a widely spread strategy in transportation and logistics, which aims for increased operational and economic efficiency, by combining cargo of different shippers and with different origins and destinations, for loading into the same vehicle for their complete or partial journeys. Consolidation-based carriers move a large and valuable part of the world trade over short, medium, long, and intercontinental distances. To achieve their economic and quality-of service objectives, carriers organize operations according to a transportation plan optimizing the resource utilization and deployment through an efficient and profitable service network addressing the estimated shipper demand. Building this transportation plan is part of the carrier tactical planning process, and involves addressing many strongly interrelated activities, marshalling several types of human and material resources. Consolidation and building a good tactical plan therefore require advanced planning methodology, and Service Network Design is the methodology of choice to address the carrier's system-wide tactical-planning challenge. The chapter presents a comprehensive overview of the general SND methodology, focusing on the modeling issues, choices, and consequences in terms of problem structure and problem solution challenges. It recalls the structure and main components of the physical and service networks of consolidation-based freight carriers, the associated tactical planning issues, and the Service Network Design methodology used to address them. It identifies and discusses in some detail the modeling the main components of consolidation-based carrier systems for tactical-planning and Service Network Design formulations. It also discusses in some depth the characteristics and modeling of three important extensions of the basic Service Network Design formulations: resource-management concerns, time characteristics of demand and services, and the explicit consideration of uncertainty. While not focusing on algorithmic developments, we mention important developments in the field.

Keywords: Service Network Design, freight transportation, consolidation, carrier service network, tactical planning.

Acknowledgements. While working on the project, the author held the UQAM Chair on Intelligent Logistics and Transportation Systems Planning, and was Adjunct Professor, Department of Computer Science and Operations Research, Université de Montréal. We gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery grant program. We also gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants.

* Corresponding author: teodorgabriel.crainic@cirrelt.net

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2024

1 Introduction

Freight transportation, and the associated flows of power units (e.g., rail engines and road tractors), vehicles, and convoys, together with the people manning them, is an important industrial, economic, and social domain, as it results from the trade exchanges generated by the needs and requirements of people, and, hence, of all public and private organizations making up the human society. It is thus widespread to describe transportation as the output of the interplay between *supply* and *demand*, within a given political, social, cultural, and economic, environment (see. e.g., Crainic et al., 2018b, and references within). To simplify the presentation, and following Bruni et al. (2024), we thus identify *shippers*, which generate the multicommodity, origin-destination demand side of the system, and *carriers* supplying the transportation and terminal resources, services, and capacity required to satisfy the demand.

Producers, traders, brokers, importers/exporters, whole- and retail buyers, sellers, and distributors of goods may be described as shippers. On the supply side, the general “carrier" term encompasses the actual modal and intermodal freight carrier organizations, which may be *private*, that is, supporting the activities and goals of the parent organization only, or, for the most part, *for-hire*, offering their services for a price/tariff to shippers. The facility (terminals, warehouses, depots, fulfillment and distribution centers, etc.) operators are grouped under the term "carrier" as well.

A special note for logistics-service provider firms, which act as intermediaries between shippers and carriers. Also known as Party Logistics (PL) firms, they generally do not own transportation means. They rather, on the one hand, act as carriers proposing transportation and logistics services to shippers, while, on the other hand, contracting with carriers to move the goods they received from their customers. It is in their carrier role that we treat them in this chapter.

One distinguishes between *dedicated* (*Full-Load*) and *consolidation-based* carriers. The former refers to the case when a loading unit, e.g., a vehicle (trailer for full-truckload motor carriers, vessel in maritime tramp shipping, barge in rivers navigation) or convoy (e.g., unit trains carrying minerals, cereals or petroleum products) is dedicated to a unique shipper demand, which pays for the complete journey. The latter combine several shipments, of different shippers with potentially different origins and destinations, for loading into the same vehicle or container for their complete or partial journeys. Such *consolidation* is a widely spread strategy in transportation and logistics, aimed to address both shipper concerns relative to shipments whose volume or value is too low to justify paying the tariffs associated with a direct, dedicated transport, and the carrier's own preoccupation of not being able to offer profitable direct service with reasonable service quality (e.g., not waiting beyond customer willingness for vehicles to fill up with other demands) in such cases. Consolidation thus should reduce the unit shipment cost and the journey time, benefiting all parties involved. Railways, Less-than-Truckload (LTL) motor carriers, shipping companies moving containers on oceans, seas, rivers, and canals, postal services and express couriers, PL firms, as well as synchromodal and City Logistics systems are prime examples of consolidation-based carriers moving a large and valuable part of the world trade over short, medium, long, and intercontinental distances.

We focus on the Operations Research methodology addressing the medium-term planning of operations and resources of for-hire consolidation-based carriers in this chapter.

Consolidation-based carriers thus aim to profitably and efficiently satisfy the requests of many shippers with the same transportation resources and services operating among the terminals of their networks. To achieve this goal, carriers organize their operations according to a transportation plan optimizing the resource utilization and deployment through a (more or less scheduled) service network that answers the expected/estimated shipper demand. To better answer their customers' demands and their own efficiency and profitability, the plan is drawn to be repetitively executed for a certain planning medium-term horizon (from a few weeks to a few months in length) for which there is a significant volume of regular and repetitive demand. The process building this transportation plan is part of the tactical planning of the carrier's activities, and is therefore sometimes called *tactical plan* in the literature.

Tactical planning involves accounting for several different but strongly interrelated activities, marshalling several types of human and material resources. These activities, taking place either in terminals or in moving between them, are characterized by not-necessarily converging individual objectives and strong, network-wide interacting impacts. Trade-offs must thus be achieved in planning the service network and schedule; First, to balance customer demand for faster and cheaper transportation, on the one hand, and the pursuit of economies of scale and profitable and efficient carrier activities, on the other hand; Second, among the various components of the carrier transportation system and operations in terms of costs, time, and resource utilization. Consolidation and building a good tactical plan therefore require advanced planning methodology.

Service Network Design (*SND*) is the methodology of choice to address these tactical-planning issues network-wide and determine the services and itineraries to operate. Closely related to Network Design, SND has been largely applied to consolidation-based freight transportation, as witnessed by a rich body of literature (reflected in this chapter and summed up in Crainic et al., 2021b). We emphasize these relations in this chapter, as well as the particular characteristics applications bring to SND.

The goal of this chapter is to present a comprehensive overview of the general SND methodology, focusing on the modeling issues, choices, and consequences in terms of problem structure and problem-solution challenges. While not focusing on algorithmic developments, we mention important developments in the field

The chapter is organized as follows. Section 2 recalls the structure and main components of the physical and service networks of consolidation-based freight carriers, the associated tactical planning issues, and the *Service Network Design* (*SND*) methodology used to address them. It also introduces the basic notation, together with the static and deterministic SND formulations. Section 3 discusses the modeling of the main components of consolidation-based carrier systems for tactical-planning SND formulations, namely, the physical system, the demand, and the service network. Sections 4, 5, and 6 discuss in some depth the characteristics and modeling of three important extensions of the basic tactical-planning process and SND formulations: time characteristics of demand and services, resource-management concerns, and the explicit consideration of uncertainty, respectively. We conclude in Section 7 with a number of research perspectives we deem important and challenging.

2 Consolidation-based Freight Transportation

We briefly recall the basic definitions and concepts of the supply facet of consolidation-based freight transportation, the physical and service networks in particular (Section 2.1), the tactical planning scope and objectives (Section 2.2), the general *Service Network Design* (*SND*) methodology and associated modeling components (Section 2.3), and the basic SND formulations (Section 2.4).

General surveys and syntheses of consolidation-based freight transportation with significant emphasis on planning and service network design may be found in, e.g., Crainic and Roy (1988); Crainic and Laporte (1997); Crainic (2000, 2003); Wieberneit (2008); Crainic and Hewitt (2021). Survey papers addressing multimodal and intermodal transport and planning, include Macharis and Bontekoning (2004); Crainic and Kim (2007); Bektaş and Crainic (2008); SteadieSeifi et al. (2014). Contributions to particular fields, mostly modal-defined, may be found in, e.g., Assad (1980); Crainic (1988); Cordeau et al. (1998); Newman et al. (2002); Ahuja et al. (2005a); Crainic (2009); Yaghini and Akhavan (2012); Chouman and Crainic (2021) for railways; Delorme et al. (1988); Bakir et al. (2021) for LTL motor carriers; Ronen (1983, 1993); Christiansen et al. (2004, 2007, 2020, 2021) for maritime transportation; Bektaş et al. (2017); Crainic et al. (2023a, 2021c); Marcucci et al. (2024) for City Logistics; Montreuil (2011); Ballot et al. (2014); Crainic and Montreuil (2016); Crainic et al. (2023c) for Physical Internet, and Giusti et al. (2019); Ambra et al. (2019, 2021) for Synchromodality.

2.1 The Carriers

Carriers providing consolidation-based services operate on infrastructure-defined single, multi, or intermodal *physical networks*, the latter term being generally used when freight packaged at origin, mostly in containers, is not handled before it is unpacked at destination. LTL motor carriers, railroads, and maritime shipping companies are identified as single-mode, when operating exclusively trucks, trains, and ships, respectively. Postal / express-courier services, container transport, as well as City Logistics, Physical Internet, and synchromodal systems often involve more than one transportation mode, the transfer of loads from one to the next taking place at intermodal terminals. Notice, however, that carriers traditionally classified as single mode often operate multi/intermodal networks. Railroads, owning motor carriers, and maritime shipping companies, owning railroads or motor carriers, illustrate this case.

The carrier physical network is thus made up of physical nodes, the *terminals*, connected by modal, e.g., highways and rail tracks, or conceptual, e.g., maritime and air corridors, links. One identifies two main categories of terminals. The largest category consists of *local/regional* terminals, where most of the demand from the corresponding regions is brought in to be transported by the carrier, and where the freight flows coming from different regions terminate their trips before being distributed to their final destinations. Rail stations, LTL regional terminals, most deep-sea and river/canal ports belong to this type. The *hubs* make up the second category, encompassing LTL breakbulks, major classification/blocking railroad yards, and large maritime ports for intermodal (container-based) traffic. While hubs act as the regional terminals for their hinterlands, their main role is to *consolidate* the flows in and out of their associated regional terminals for efficient long-haul transportation and economies of scale. Terminals may be owned/managed by and dedicated to the carrier, e.g., railroad yards and LTL breakbulk terminals, or may be shared by several carriers irrespective of ownership and management, e.g., maritime ports and terminals, intermodal terminals, and airports. Inter-terminal links may also be proprietary (but may still be used by other carriers for a fee), e.g., rail tracks in North America, or shared, e.g., rail tracks in Europe and roads and highways mostly everywhere.

The multi-commodity, *origin-destination* (*OD*) *demand* is defined, on that same network, as a quantity (weight and volume) of freight to be moved between specific origin and destination terminals. Many other attributes may be associated to each demand, according to the particular context, including shipper identity and demand type (e.g., priority, express, regular delivery), timing constraints (e.g., availability at origin and due-date at destination), fare and penalties for missing the agreed-upon service-quality targets, particular commodity (product) with related physical characteristics and requirements in terms of vehicle type (e.g., refrigerated, multi-platform for containers or vehicles, etc.), and so on.

Consolidation-transportation carriers aim to satisfy this demand by organizing their operations into so-called *hub-and-spoke service networks*, schematically illustrated in Figure 1. Each *service* is defined by a (modal) physical route between a pair of origin and destination terminals (the full, dash, and point-dash line patterns in the figure illustrate three different main modes, while the dotted line patterns illustrate feeder services), a possible sequence of intermediate stops on that route where loads may be picked up / consolidated / dropped off and resources may be handled (the dash-point service between hubs A and B, with a stop) at hub C), a schedule indicating more or less precisely arrival and departure times at/from the concerned terminals, as well as particular equipment, operations, and economic characteristics.

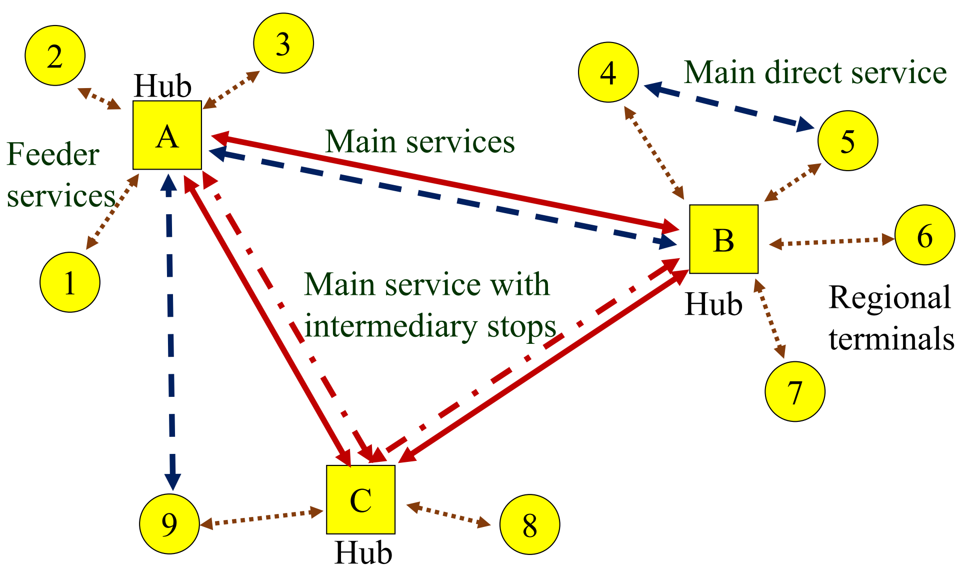

Figure 1: Hub-and-spoke service network

Using the service network, carriers thus first move low-volume loads available at a regional terminal to a hub, by using so-called *feeder* services. At hubs, loads are sorted (*classified* is the term used in several settings, e.g., freight railways) and consolidated into larger flows, which are routed to other hubs by high-frequency, high-capacity services. Loads may thus go through more than one intermediary hub before reaching the regional-terminal destination, being transferred from one service to another or undergoing re-classification and re-consolidation. Notice that, when the level or value of demand justifies it, high-frequency, high-capacity services may be run between a hub and a regional terminal or between two regional terminals. Notice also that, more than one service, of possibly different modes, may be operated between consolidation and regional terminals. Once at the last hub on their itineraries, loads are unloaded, possibly sorted, and loaded on feeder services to be moved to their destination regional terminal, to be distributed from there to their final destinations.

It is noteworthy that railways generally implement a more complex double consolidation policy, which groups (classifies) cars into *blocks*, which are then consolidated to make up *trains*. More precisely, loaded and empty cars, with different origins and destinations, being present simultaneously in the same terminal, are sorted and grouped into a block, which is then moved as *a single unit* by a series of trains until its destination, where it is broken down, the cars being either delivered to their final consignees or classified again into new blocks. More than one reclassification may make up the itinerary of a given demand flow. Blocks, on the other hand, travel on a series of train services, being simply *transferred* (*switched*) from one train to another at intermediary stops on their routes.

A hub-and-spoke network concentrates the multi-commodity flows and allows a much higher frequency of service for the consolidated demand loads, while providing a more efficient utilization of resources, economies of scale for the carrier, and lower tariffs for the customers. The drawbacks of this type of organization are possibly increased delays for demand due to longer routes and more time spent going through terminals, which play a role significantly broader than the loading/unloading of freight within consolidation-based transportation systems. Vehicle and freight classification and consolidation, convoy making-up and dismantling, and vehicle and freight transfer between services are all time, cost, and resource-consuming operations performed in terminals.

The adequate, hopefully optimal in Operations Research terms, design of the service network and planning of its operations and resource utilization, is required to avoid such pitfalls and reap the full benefits of consolidation-based transportation. This is the scope of carrier tactical planning described next.

2.2 Tactical Planning

Carriers need to be profitable while answering the demand of shippers requiring low tariffs and high quality, timely service. Consolidation offers the operation environment to fulfill these goals, but also raises several challenges.

First, on the demand side, services cannot be tailored to address shipper demands individually, as the different loads of several shippers are grouped within the same vehicles and convoys. Hence, the service network has to address globally the requirements of a large group of shippers, which contracted or which may contract with the carrier for the period considered. This has implications for the planning of the service network, including the so-called *planning* *horizon* (i.e., duration of validity), topology (i.e, where and what services to propose and operate), timeline (i.e., when to operate services), and performance measures (e.g., cost, efficiency, and quality of service).

Material and human resources are needed to operate services, their management significantly impacting carrier service efficiency and profitability. Resources are costly, however, and their availability is continuously and increasingly limited, due to high acquisition and operation costs for moving and in-terminal equipment, and the shortage (and salaries) of manpower. Then, as the cost of operating a service is greatly dependent upon the costs of vehicles, power units, and crews used for transport and terminal work, higher economies of scale may be achieved by assigning the most appropriate resources to each service to fill it well given its planned load. On the other hand, idling resources are unproductive and, thus, one aims for seamless multi-service routes with inter-service transfers at terminals.

The supply-side challenge therefore is to plan service networks and operations, which are efficient from the point of view of using the carrier's resources (own, rented, or obtained through outsourcing part of operations) and, thus, achieve the best trade-off between shipper requirements and expectations, and carrier pursuit of profitable efficiency. *Tactical planning* aims to answer this challenge.

The scope and goal of tactical planning is to build a *transportation plan*, often together with an associated schedule, to mitigate the drawbacks of consolidation, satisfy customer demand and service-quality requirements, and operate profitably and efficiently. It addresses the system-wide planning of operations and resources, deciding on the selection and scheduling of services, the transfer and consolidation activities of freight and vehicles in terminals (as well as the convoy makeup and dismantling for rail, road, and barge trains), the assignment and management of resources to support the selected services, and the routing of freight of each particular demand through the resulting service network. The goal is cost-efficient operation together with timely and reliable delivery of demand according to customer specifications and the service-quality targets of the carrier.

Such planning problems are difficult due to the strong interactions among system components and decisions, and the corresponding trade-offs between operating costs and service levels that need to be achieved. Two examples to illustrate these challenges and necessary trade-offs. First, increasing the number of services operated during a certain time interval between two terminals improves customer service but may decrease the availability of resources for other services, as well as increase congestion in terminals, and on certain infrastructure links such as rail tracks, thus increasing costs and deteriorating customer service. Second, consider the strategies, proper to consolidation transportation, based on routing of freight through intermediate terminals, where incoming loads are re-classified and consolidated before transfer to the next service. Such strategies generally result in better equipment utilization and lower waiting times at the original terminals; hence, in a more rapid service for the customer. The same strategies would also result, however, in additional unloading, consolidation, and loading operations, involving larger delays and creating higher congestion levels at terminals. Transfers may also decrease the shipment delivery reliability. The alternative, offering more direct and frequent services, would imply faster and more reliable service for the corresponding traffic and a decrease in the level of congestion at some terminals, but at the expense of additional resources and, thus, higher costs for the system and tariffs for the shippers.

Tactical planning is performed for a so-called medium-term planning horizon, which could extend from a few (e.g., LTL motor carriers) to six or more (but twelve at maximum) months, and which we call *season*. It builds the transportation plan considering the so-called *regular demand.* One generally finds in this category shippers that are strongly believed to bring business on a regular basis for the coming season. This prediction, which may or may not follow from formal forecasting methods, is based on a combination of long-term contracts, informal understanding with long-standing and trustful customers, and market estimation by sales and customer-relation personnel. In terms of volume, regular demand is expected to make up a good part, e.g., 75% - 80% of the pick demand to be serviced on a “normal" operating day. In terms of consistency, demand, and, hence, service, is expected to be repetitive according to a certain pattern, e.g., every day or every week. The plan produced by the tactical-planning process is thus for a given time duration, called *schedule length*, and is to be applied repetitively for the duration of the season. Of course, particular service networks may be build for specific moments, e.g., for week days different from weekends, and be combined to form the repetitive plan and schedule length.

Service Network Design (*SND*) is the methodology of choice to support tactical planning of consolidation-based carriers, addressing these issues and determining the services and itineraries to operate. The SND methodology is introduced in the next subsection and is further detailed in Sections 3 - 6.

2.3 Service Network Design

A service network design model for tactical planning of a consolidation-based carrier yields the transportation plan specifying operations for a given schedule length, to be repetitively applied for the next season.

Its input follows the transportation-planning structure and includes demand and supply components. The former gives the origin-destination matrix (or matrices, when several products are considered) of the quantities of freight to be moved between particular pairs of origin and destination terminals. The latter specifies the physical network, the set of potential services, and the available resources, out of which the service network is to be built to answer this demand. The physical and operational attributes of demand and supply elements, as well as the rules associating them and specifying how the system is to be operated, are also part of the input of SND models. Section 3 discusses the modeling of these elements and rules within various carrier-planning contexts.

Through its decision variables, a SND model represents and integrates two major sets of decisions, one related to the design of the service network, and one associated to the utilization of that network to service demand. The main *design decision* concerns building the service network by selecting the services to be operated out of the set of potential services. When the management of resources is explicitly accounted for within the tactical-planning decision process, decisions relative to the association of resources to services and to the routing and handling of those resources are also part of the design decision set.

The second major set of decisions concerns how demand for transportation is satisfied using the designed service network. Such *utilization* or *flow* decisions concern, for each individual demand, the *itinerary*, i.e., the sequence of services, terminals, and terminal operations (loading, unloading, inter-service transfer, classification and consolidation), used to move the corresponding flow. Several itineraries may be used simultaneously for a given demand, when its shipment may be split among several service paths between the respective origin and destination terminals.

The objective function of an SND formulation reflects the economics of the carrier, as well as, increasingly, its customer-service and societal concerns and objectives. Carriers aim to maximize their profit, generally defined as revenue minus cost. Most tactical-planning applications assume the set of shippers and corresponding commercial relations to be known. Hence, the expected/estimated revenue is also assumed to be known and fixed for the tactical planning horizon. Consequently, the transportation plan aims for the minimization of the total operating costs, which also reflects the traditional carrier and shipper objectives to "get there fast at the lowest possible cost". Profit maximization is considered, however, when revenue-management policies are applied by the carrier and are thus reflected in the objective function of the SND model (see Section **??**) .

Increasingly, however, shippers expect low rates for high-quality service, measured, in particular, by speed and reliability in service and delivery times. Simultaneously, carriers are increasingly sensitive to social and legal pressure regarding energy consumption and environmental impacts. Service performance measures reflecting these expectations and concerns are modeled, in most cases, by delays incurred by freight and vehicles or by the respect of predefined performance targets These measures are then added to the objective function of the SND optimization formulation, yielding a generalized cost function that captures the trade-offs between operating costs and service quality. We discuss these issues in Section 3.3.

One generally classifies SND problems and models relative to two dimensions, time and uncertainty. Along the *time* dimension, one finds *static* and *time-dependent* (the term "time-sensitive" is also used) problem settings and formulations. The former, discussed in this section and the next one, assumes that neither demand, nor any other problem characteristic varies during the schedule length and the planning horizon considered. Time-dependent settings, Section 4, include an explicit or implicit representation of demand and activities in time, for the schedule-length duration, and target the selection of *scheduled* services to support decisions related to *when* services and freight leave and arrive at terminals on their routes and itineraries, respectively. These schedule-length-specific time-dependent variations are assumed to be repeatedly present during the planning horizon, hence, the scheduled tactical plan built for that schedule length is to be applied repeatedly for the planning horizon.

The second major dimension concerns uncertainty issues. Most contributions in the literature assume known values for the system parameters, both on the demand and the supply facets, which will not change for the planning-horizon duration. This translates in fixed figures, which may come from historical and field-knowledge data (one finds a lot of average measures) or single-point estimations, when forecasting methods are used. The problem settings and formulations discussed in Sections 2 - 5 belong to this class. The future is uncertain, however, and the explicit representation of uncertainty of various system parameters, together with the corresponding plan-adjustment actions and costs, when new information becomes available, is increasingly part of the tactical-planning and SND literature. We discuss these issues, challenges, and modeling avenues in Section 6.

2.4 The Fundamental SND Models

We introduce the basic notation and the arc- and path-based SND formulations addressing the classic, static deterministic, single-mode, direct-service, linear-cost problem setting. Here as in the rest of the chapter, we follow the notation of Crainic and Hewitt (2021).

The fundamental system and model components of all SND problem settings and formulations, discussed in some depth in Section 3, are:

Physical network, $\mathscr{G}^{\mathrm{PH}} = (\mathscr{N}^{\mathrm{PH}}, \mathscr{A}^{\mathrm{PH}})$, where $\mathscr{N}^{\mathrm{PH}}$ stands for the set of facilities, hubs and regional terminals, connected by the physical or conceptual links of set $\mathscr{A}^{\mathrm{PH}}$ (Section 3.1);

Demand for transportation of a set $\mathscr{K}$ of OD commodities, each $k \in \mathscr{K}$ requiring to move a quantity of freight vol_k from its origin $O(k)$ to its destination $D(k)$ (Section 3.2);

Service network $\mathscr{G} = (\mathscr{N}, \mathscr{A})$, defined based on the physical nodes of the system and the set of potential services Σ, within the context of the carrier resources, operation rules, and economics and service goals (Section 3.3).

Resources, and their management-concerned assignment to services, enrich this basic setting and are discussed in Section 5. Time-related characterization of system elements and decisions are the topic of Section 4.

In the basic setting, $\mathscr{N} = \mathscr{N}^{\mathrm{PH}}$ and $\mathscr{A} = \Sigma$, with $\sigma(a) \in \Sigma$ and $a(\sigma) \in \mathscr{A}$ denoting the service-arc association. Each service $\sigma \in \Sigma$ is characterized by a fixed cost Σ, incurred when selecting and operating it, a unit freight-transportation cost c_σ or $c^k_\sigma, k \in \mathscr{K}$, when commodity characteristics are relevant, and a capacity u_σ, representing the total volume of freight the service may load and haul; particular, commodity-specific capacities $u^k_\sigma, k \in \mathscr{K}$, are defined when relevant (e.g., hazardous material-loaded railcars on general trains). The cost and capacity figures are thus inherited by the corresponding arc $a \in \mathscr{A}$, i..e., $c^k_a = c^k_\sigma$ ($c_a = c_\sigma$) and $u_a = u_\sigma$ ($u^k_a = u^k_\sigma$).

Two sets of decision variables are found in all SND settings, *design* or service *selection* for inclusion in the service network and tactical plan to be executed repeatetly diring the planning horizon, and *flow* or *utilization* of the service network by the commodity demand. The design SND decision variables in the basic setting model the selection of services through binary variables $y_\sigma \in \{0,1\}, \sigma \in \Sigma$. The flow decisions variables are generally defined on the arcs of the service network $\mathscr{G}$, which is the current case on $a \in \mathscr{A}$, and take the form $x^k_a \geq 0$, $a \in \mathscr{A}, k \in \mathscr{K}$, prescribing the amount of commodity k that travels on the arc, i.e., on service $a(\sigma)$.

Formally, then, the basic, linear-cost, SND formulation seeks to

$$\min \sum_{\sigma\in\Sigma} f_\sigma y_\sigma + \sum_{k\in\mathscr{K}} \sum_{a\in\mathscr{A}} c_a^k x_a^k \tag{1}$$

$$\text{s.t.} \sum_{a\in\mathscr{A}_\eta^+} x_a^k - \sum_{a\in\mathscr{A}_\eta^-} x_a^k = d^k \qquad \eta\in\mathscr{N}, k\in\mathscr{K}, \tag{2}$$

$$\sum_{k\in\mathscr{K}} x_a^k \leq u_a y_{\sigma(a)}, \qquad a\in\mathscr{A}, \tag{3}$$

$$x_a^k \leq u_a^k y_{\sigma(a)}, \qquad a\in\mathscr{A}, k\in\mathscr{K}, \tag{4}$$

$$y_\sigma \in \{0,1\}, \qquad \sigma\in\Sigma, \tag{5}$$

$$x_a^k \geq 0, \qquad a\in\mathscr{A}, k\in\mathscr{K}. \tag{6}$$

where $\mathscr{A}_\eta^+ = \{(\eta,\eta')\in\mathscr{A}\}$ and $\mathscr{A}_\eta^- = \{(\eta',\eta)\in\mathscr{A}\}$ define the sets of incoming and outgoing arcs, for node $\eta\in\mathscr{N}$, respectively, while $d^k = \text{vol}_k$ at the demand origin $\eta = O(k)$, equals $-\text{vol}_k$ at the destination node of the demand $\eta = D(k)$, and zero at all other nodes.

The objective of the SND minimizes the total cost of operating the system, computed as the sum of the fixed costs associated with selecting the service network and the variable cost of transporting commodities using the selected services. Equations (2) are often referred to as *flow-balance* constraints and ensure that all of a commodity's demand departs from its origin (the first case), arrives at its destination (the second case), and departs any other locations at which it arrives (the third case). The expression on the left-hand side of the *linking* constraints (3) computes the total flow traveling on arc $a\in\mathscr{A}$, whereas the expression on the right-hand side gives the global arc capacity provided by the corresponding service (selected or not). The commodity-disaggregated linking constrains are given by (4). Constraints (21) and (6) define the variable domains.

The flow variables x describe how the demand loads are moving through the selected service network. Clearly, the flow of each demand follows one or several paths from its origin to its destination. An equivalent formulation explicitly determines these paths.

Let $\Pi^k, k\in\mathscr{K}$, identify the set of possible paths of commodity k, on the potential service network (all potential services and flow-terminal operations). Following classic network notation, let set $\mathscr{A}_\pi^k$ hold the sequence of arcs $a\in\mathscr{A}$ (services in Σ_π^k) making up the itinerary $\pi\in\Pi^k$, $\delta_a^{\pi k}$ be the Kronecker delta defining this path, i.e., $\delta_a^{\pi k} = 1$ when $a\in\mathscr{A}_\pi^k$, 0, otherwise, and the unit path flow costs $c_\pi^k = \sum_{a\in\mathscr{A}_\pi^k} c_a^k, k\in\mathscr{K}$.

Define the *path-flow* decision variable h_π^k as the amount of commodity $k\in\mathscr{K}$ moved on its path $\pi \in \Pi^k$. The basic, linear cost, path-based SND formulation then becomes

$$\min \sum_{\sigma\in\Sigma} f_\sigma y_\sigma + \sum_{k\in\mathscr{K}} \sum_{\pi\in\Pi^k} c^k_\pi h^k_\pi \tag{7}$$

$$\text{s.t.} \sum_{\pi\in\Pi^k} h^k_\pi = d^k, \qquad k \in \mathscr{K}, \tag{8}$$

$$\sum_{k\in\mathscr{K}} h^k_\pi \le u_a y_{\sigma(a)}, \qquad a \in \mathscr{A}, \tag{9}$$

$$h^k_\pi \le u^k_a y_{\sigma(a)}, \qquad a \in \mathscr{A}, k \in \mathscr{K}, \tag{10}$$

$$y_\sigma \in \{0,1\}, \qquad \sigma \in \Sigma, \tag{11}$$

$$h^k_\pi \ge 0, \qquad \pi \in \Pi^k, k \in \mathscr{K}. \tag{12}$$

Recall that, the network-design arc and path-based formulations are equivalent, that is, they yield the same service network and objective-function value, with $x^k_a = \sum_{\pi\in\Pi^k} \delta^{\pi k}_a h^k_\pi, a \in \mathscr{A}, k \in \mathscr{K}$ (Crainic et al., 2021a). This holds for the SND formlation above, as well as for the straightforward generalization to the multi-stop service case.

The particular characteristics of the various carriers, modes, and operation policies yield a number of extensions to this basic formulation, namely

Multiple design layers. While many SND models address problem settings considering a single level of consolidation (shipments into a vehicle or container), many other address planning issues for modes involving multiple layers of consolidation (e.g., the car-to-block-to-train strategy of railway transportation). Similar modeling requirements arise when resource-management concerns are part of tactical planning (Section 5), as well as in applications to motor-carrier platooning for long-haul movements (Scherr et al., 2019; Albinski et al., 2020; Ammann et al., 2024).

Such *multi-layer* service networks (Crainic et al., 2022; Crainic, 2024) display the particular characteristic of an arc in a given decision layer being defined with respect to a set of arcs, often making up a path or a cycle, in another decision layer. The railway case illustrates the concept as each potential block is defined in the block layer, in terms of the path of service arcs that would transport it, if selected in the service layer (Zhu et al., 2014; Chouman and Crainic, 2021). Such interwoven definitions imply several connectivity relations and requirements in terms of both design and flow-distribution decisions, yielding rich *Multi-layer Network Design* (*MLND*) formulations raising challenging algorithmic issues (Crainic, 2024).

To define the basic multicommodity, fixed-cost, capacitated MLND formulation, let $\mathscr{L}$ be the set of layers of multi-layer network $\mathscr{G} = \bigcup_{l\in\mathscr{L}}\{\mathscr{G}_l\}$, where $\mathscr{G}_l = (\mathscr{N}_l, \mathscr{A}_l)$ is the network on layer $l \in \mathscr{L}$, with $\mathscr{N}_l$ and $\mathscr{A}_l$ the corresponding sets of nodes and arcs. Let $l, l' \in \mathscr{L}$ be a couple of *(supporting, supported*) layers of $\mathscr{G}$ (e.g., (train, block)), coupled by an arc definition specifying how an arc in *supported* layer l' is related to a subset of *supporting* arcs in layer l (e.g., the supporting train service arcs form the path defining the supported block). For simplicity of presentation, we assume that all arcs in $\mathscr{G}$ are design arcs, and that a single set of OD demands $\mathscr{K}$ is defined on a given layer (notice that, the flows on that layer are be projected on the layers associated with it through the arc definitions). The rest of the SND already-defined notation applies in this case as well (adjusted with the appropriate layer index), including for the decision variables $y_{al} = 1$ if arc $a \in \mathscr{A}_l$ of layer l is selected, 0, otherwise; and x^k_{al} indicating the quantity of demand $k \in \mathscr{K}$ assigned to arc a of layer l. The basic MLND formulation then becomes, where relations (17) stand for the sets of constraints corresponding to the design, flow, or attribute connectivity requirements proper to the multi-layer network design application at hand:

$$\min \sum_{l \in \mathscr{L}} \left\{ \sum_{a \in \mathscr{A}_l} f_{al} y_{al} + \sum_{k \in \mathscr{K}} \sum_{a \in \mathscr{A}_l} c^k_{al} x^k_{al} \right\} \tag{13}$$

$$\text{s.t.} \sum_{a \in \mathscr{A}^+_{\eta l}} x^k_{al} - \sum_{a \in \mathscr{A}^-_{\eta l}} x^k_{al} = d^k_\eta, \qquad \eta \in \mathscr{N}_l, k \in \mathscr{K}, l \in \mathscr{L}, \tag{14}$$

$$\sum_{k \in \mathscr{K}} x^k_{al} \leq u_{al} y_{al}, \qquad a \in \mathscr{A}_l, \tag{15}$$

$$x^k_{al} \leq u^k_{al} y_{al}, \qquad a \in \mathscr{A}_l, k \in \mathscr{K}, \tag{16}$$

$$(\mathbf{y}, \mathbf{x}) \in (\mathscr{Y}, \mathscr{X})_{ll'}, \qquad (l, l') \in \mathscr{C}, l \in \mathscr{L}, \tag{17}$$

$$y_{al} = 1 \in \{0, 1\}, x^k_{al} \geq 0, \qquad a \in \mathscr{A}_l, k \in \mathscr{K}. \tag{18}$$

Service frequency. Independently of operating according to fixed schedules or not, a service may have several departures during a given time interval. Such cases are modeled by defining non-negative integer service-selection decision variables $y_\sigma \in \mathbb{Z}_+, \sigma \in \Sigma$;

Service capacity feasibility. Including explicitly capacity limitations Operations Research models may raise feasibility issues. These are disturbing, both from a computational point of view and because, in practice, there is “always" a feasible solution, even if quite costly, e.g., by calling on ad-hoc capacity provided by additional vehicles or outsourcing part of the demand transportation, and paying the additional costs. The simplest approach to address this issue is to include *dummy arcs* t $a^k = (O(k), D(k))$ between the origin and destination of each commodity $k \in \mathscr{K}$, with no capacity restrictions and appropriately high unit cost c_{a^k}. The associated *slack-flow variable* ς^k then captures the volume of the demand unfulfilled by the capacity of the selected services, and takes care of the feasibility issue. The flow conservation constraints (2) then become

$$\sum_{a \in \mathscr{A}^+_\eta} x^k_a - \sum_{a \in \mathscr{A}^-_\eta} x^k_a = \begin{cases} \text{vol}_k - \varsigma^k, & \text{if } \eta = O(k), \\ -\text{vol}_k + \varsigma^k, & \text{if } \eta = D(k), \\ 0, & \text{otherwise}, \end{cases} \quad \eta \in \mathscr{N}, k \in \mathscr{K}, \tag{19}$$

and the term $\sum_{k \in \mathscr{K}} c_{a^k} \varsigma^k$ is added to the objective function. (These modifications are implicit when the set of dummy arcs a^k is included in $\mathscr{A}$.)

Demand distribution. The previous formulations address the case where the volume of any particular OD demand may be *split* between several paths within the service network. While this corresponds to a very large range of applications, one equally finds many situations where the freight of an OD demand must travel together on a single path. This is achieved by modifying the definitions of the flow variables

- x_a^k equals 1 if commodity $k \in \mathscr{K}$ travels on the arc $a \in \mathscr{A}$, i.e., on service $a(\sigma)$, and 0, otherwise;
- h_π^k equals 1 if commodity $k \in \mathscr{K}$ is moved on its path $\pi \in \Pi^k$, and 0, otherwise .

The arc (path) formulation is then modified by multiplying x_a^k (h_π^k, respectively) by d^k in the objective function (1) ((7)) and constraints (2) - (4) ((8) - (10)), and by changing the domain restrictions of the flow variables (6) to $x_a^k \in \{0,1\}$ ((12) to $h_\pi^k \in \{0,1\}$).

Objective function. While the linear-cost formulation (1)represents a very broad set of issues and problem settings, other cases required more elaborate objective-function formulations. Among these cases, further discussed in Section 3.3, one may cite modeling of the total cost associated to several more-or-less simultaneous service departures through non-continuously differentiable functions (e.g., Croxton et al., 2003a,b, 2007), and the explicit representation of congestion phenomena, in terminals or on the infrastructure, and penalties for violating capacity restrictions or service-quality targets through non-linear functions (e.g. Crainic et al., 1984; Crainic and Rousseau, 1986).

On line with the two-facet decision-making characteristics of the problem at hand, namely design and flow distribution, we represent such cases by defining

- $\phi_\sigma(y), \sigma \in \Sigma$: Fixed cost of selecting service σ given the selected services y;
- $\varphi_{ak}(y,x), k \in \mathscr{K}$: Unit-transportation cost of commodity $k \in \mathscr{K}$ on arc $a \in \mathscr{A}$, given the selected services y and flow distribution x;

Particular system features, such as limited terminal capacity to handle vehicles and freight, freight-vehicle compatibility restrictions, and budgetary limits, are other encountered in carrier-planning applications. We discuss a number of them further in the chapter, for now representing them through a general set Ψ linking the decision variables (y,x) and restricting their domains.

The general basic SND formulation then takes the form

$$\min \sum_{\sigma \in \Sigma} \phi_\sigma(y) y_\sigma + \sum_{k \in \mathscr{K}} \sum_{a \in \mathscr{A}} \varphi_{ak}(y,x) x_a^k \tag{20}$$

subject to constraints (2) - (4) and (6), plus

$$y_\sigma \in \mathbb{Z}_+, \quad \sigma \in \Sigma, \tag{21}$$

$$(y,x) \in \Psi. \tag{22}$$

These basic formulations are found in many contributions in the literature targeting freight-transportation planning issues, as synthesized in the papers and chapters indicated at the start of this section.

These formulations also emphasize the network-design nature of the SND formulations. Indeed, any SND model may be cast as a ND formulation on an appropriately-defined network. Thus, model (1) - (6) corresponds to the multi-commodity, fixed-cost, capacitated network design problem (see Crainic et al., 2021a, for a state-of-the-art presentation) defined on a network with $\mathscr{A} = \Sigma$. Moreover, the general formulation (20) - (22) is the same as the general network design model defined in the seminal Magnanti and Wong (1984) paper, where the authors also showed that ND englobes Minimal Spanning Tree, Shortest Path, Traveling Salesman, Vehicle Routing, and Facility Location problems as special cases.

3 Systems & Components

This section is dedicated to going somewhat deeper in modeling the three main components of a SND formulation of a tactical-planning problem for consolidation-based freight transportation, namely, the physical system, Section 3.1, the demand, Section 3.2, and the potential service network, Section 3.3.

3.1 Physical system

Planning is performed and SND models are formulated on a network representation of the physical infrastructure on which the consolidation-based freight carrier operates. Modeling this *physical network* $\mathscr{G}^{\mathrm{PH}} = (\mathscr{N}^{\mathrm{PH}}, \mathscr{A}^{\mathrm{PH}})$ for medium-term, tactical planning means representing the infrastructure, activity, and performance of the terminals and inter-terminal connections at a level relevant for planning.

Terminals = Network nodes $\eta \in \mathscr{N}^{\mathbf{PH}}$. Only major hubs and regional terminals, making up the core activity network, are thus generally included. The other terminals, e.g., the small rail stations on feeder lines and the small ports services by coastal-navigation ships out of major ports, are generally represented through the freight originating and terminating there only, which is aggregated and assigned to the regional terminal or hub to which the feeder line is attached.

The included terminals are thus, generally, major infrastructure facilities equipped and maned to perform many types of handling operations on freight and vehicles. For tactical planning purposes, these are represented mostly through their economic and efficiency performance measures related to the equipment and its handling capability. Economic measures means costs, which can be commodity- or mode-specific, or both. while efficiency is often measured either through the throughput, e.g., volumes processed or vehicles and convoys services, or the associated duration / delay. One or several *capacity* measures are generally defined to represent the terminal operational capability. Measured for a given time duration appropriate for the planning problem at hand, e.g., a day, these measures may refer to the volume of freight, commodity-specific or no, handled, and the number of vehicles or convoys serviced. Often, a single cost or capacity measure is defined for the terminal, aggregating in an unique figure the average (usually, sometimes the variance is also given) terminal performance. Operation-specific measures are defined for particular activities with particular impacts. Thus, for example, one may find four measures for classification (or marshalling) rail yards, standing for: 1) aggregation measures of the reception, inspection, and departure activities (including loading/unloading when appropriate); 2) transfer of railcars or blocks from one train to another; 3) classification/consolidation of railcars into blocks; and 4) accumulation of railcars to form the selected block before it is attached to the departing train (e.g. Crainic et al., 1984).

The physical networks of most SND models include a single node for each terminal present in the problem setting, the performance measures and capacities being uniquely associated to it . More complex representations may be used to account for particular operations or terminal characteristics, usually taking the form of mini networks. The simplest mini network terminal representation involves two nodes, capturing the incoming and outgoing traffic, respectively, the link connecting them capturing the performance measures and characteristics of the respective terminal (e.g., Zhu et al., 2014). Such a representation provides a more refined modeling of particular trminal activities and performance measures, e.g., different traffic directions and capacities.

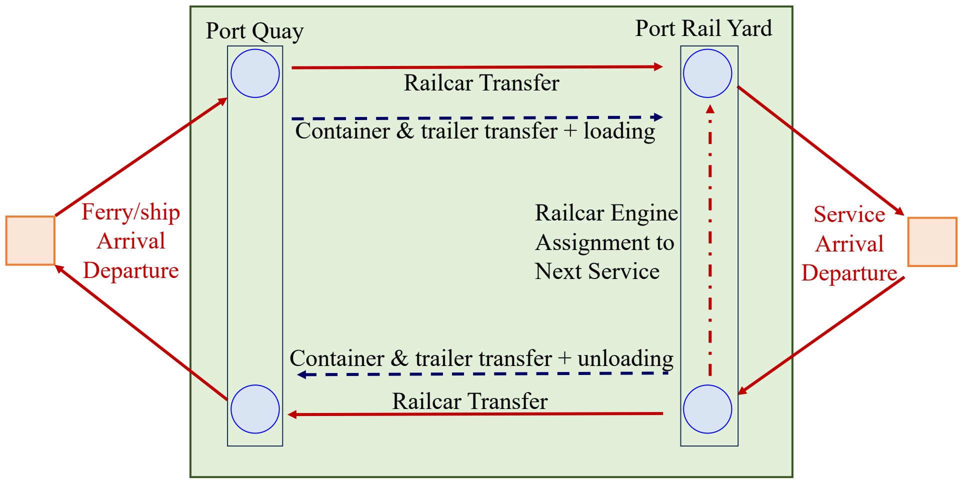

Figure 2: Mini-network terminal representation

Larger networks may capture a richer set of activities and measures. Figure 2 illustrates such structures through the representation of an intermodal terminal linked, on one side to a maritime network of ferries or roll-on/roll-off ships and, on the other side, to a rail network (inspired by Andersen et al., 2009b). Several types of nodes make up this network: two representing the navigation and rail networks, respectively (notice that, these are indicative only in the illustration; more detailed networks appear in actual applications), two standing for the port quay, and two for the rail yard of the port. These last four nodes, together with the links connecting them, make up the actual intermodal terminal (which is often, but not always, within the port compound). The links between the maritime and the port-quay networks capture the loading and unloading of both rail flat cars carrying containers and trailers, and individual containers or trailers. The former are moved on the railcar-transfer arcs towards/from the port-rail yard. The unloading from the ship, transfer, and subsequent loading on appropriate railcars of individual containers or trailers take place on the railcar transfer arcs. Similar activities bring individual containers and trailer to the ship. The link connecting the two port-rail yard nodes stands for the situation when the railcars are brought to the port and taken out of it by the railway's own engines, in-bounding ones needing to turn around and be assigned to outgoing movements. Finally, the two links connecting the port and the railroad networks stand for the blocking and unblocking of the railcars in port and their transfer to and from the rail network, respectively.

Particular cost and capacity functions or values may then be assigned to each of the particular arcs of terminal representation. Notice that, container terminals accommodating long-haul maritime liners may be represented in this way by discarding the railcar transfer arcs. Similarly, roll-on/roll-off ship terminals may include the railcar transfer arcs only.

It is worth recalling that, commodity origins and destinations are defined at regional terminals (or hubs when those act as regional terminals for their hinterlands). Hence, the appropriate node must be selected when a mini network is used for terminal modeling, e.g., in a two-node representation, the outgoing-traffic node for the origin of demand, and the incoming one for the destination.

Moving among terminals = Network arcs $a \in \mathscr{A}^{\mathbf{PH}}$. In its most general form, an arc of the tactical-planning physical network stands for the possibility to move directly, i.e., without passing by any other terminal in the network model, between the two terminals represented by the two nodes defining it. It thus represents a path in the physical or conceptual (when air or maritime networks are involved) network on which the carrier operates. Arcs in mini networks modeling particular terminals follow the same rule, but may stand more for specific operations (e.g., railcar classification), rather than actual movements.

Physical network arcs are usually *directed*, providing the means to represent direction-specific characteristics, measures, and flows. They are also *modal*, *unimodal*, in fact, that is, each arc is characterized by a specific transportation mode. Parallel arcs may be defined between two adjacent nodes, when more than one modal connection exists between the two corresponding terminals.

“Mode" is a very general term in transportation and logistics, referring to different concepts and definitions according to the topic at hand. Thus, in a very fundamental way, mode may refer to the nature support of transportation, e.g., land, water, air, and space, while, it may also refer in a very detailed way to the specifics of a transportation system, e.g., the combination of rail-track gauge (narrow, metric, or imperial), traction type (diesel or electric), and authority (state or country) defining rail modes for evaluation and planning analyses (e.g., Crainic et al., 1990a,b; Guélat et al., 1990; Crainic and Florian, 2008). Most SND problem settings refer to a “classical" definition of modes, based on a high-level combination of elements such as nature, infrastructure, and vehicle/traction, in particular

Less-than-Truckload (*LTL*) trucking on roads, particularly in inter-urban settings, with an increasingly larger array of road-based modes in cities, e.g., people-driven or autonomous electric or hydrogen (or mixed) powered vans, cargo bikes, and robots;

***Rail*,** particularly the *general* train services, performing consolidation-based rail transportation for mostly all categories of goods, and *intermodal* services, dedicated to moving containerized cargo;

***Navigation*,** with the largest part of the literature dedicated to maritime transport of containers performed by liner ships; One notices an increased interest in barge river & canal navigation and the transportation of containers;

***Air*,** with relatively few contributions related to service network design, most of which target the air mode as part of the multimodal network of express-courier firms (contributions addressing the air-cargo industry are mostly directed toward revenue-management issues).

To conclude, from a carrier point of view, most transport is unimodal and, hence, such is the physical network representation. From a shipper point of view (including intermediaries and carriers acting as such, as well as planners/analysts at the level of a regional/national/international transport system study), however, most freight movements are *multimodal* involving two or more modes, cargo being transferred from one mode to the next at *intermodal* terminals. Multimodal becomes *intermodal* transportation when cargo is packaged in such a way that it is not touched in transport nor when transferred. In its most general accepted definition, cargo is thus assumed to be packaged in containers, and many assimilate intermodal and container-based transportation. While this definition covers a good part of relevant transportation, a more general definition specifies that cargo has to be loaded into a “loading unit", which may be a container, but could also be a trailer, a swap body, and even boxes grouped on a pallet and tightly wrapped. Moreover, North American railroads introduced the term intermodal rail to designate their specific operations of handling and moving containers, often managed by particular administrative divisions of the railroad. We touch on these issues in Sections 3.2 and 5.

Several attributes are generally associated to each arc of $\mathscr{G}^{\mathrm{PH}}$, indicating the physical, material, and operational characteristics and limits of the infrastructure path defining it. The nature of each attribute determines how it is computed. *Length* is thus equal to the sum of the lengths of the links of the infrastructure path, while the *capacity* is computed as the minimum of the corresponding capacities. When defined, *time* is also computed as a sum, including, when relevant, the time required by technical activities at facilities (e.g., inspection or crew changes), not accounted for in the duration of the planned stops at intermediary terminals. It is noteworthy, however, that traveling time is dependent upon the characteristics (e.g., speed and priority) of each service and, thus, time is more often associated to the service network rather than the physical one.

The capacity is generally measured in numbers of vehicles or convoys, of the respective mode, that can pass through or be processed by the relevant infrastructure during a given time length (e.g., hour, day, or week, depending on the degree of aggregation used in the problem setting). The capacity may sometimes be also measured in length limiting, e.g., the length of trains that may used waiting side-line tracks when meeting or being overtaken by other trains, and the total length of the ships that can moor simultaneously at a given port quay. A number of other measures may be encountered, particularly for the arcs of mini-networks modeling terminals, e.g., the weight or volume that may be processed or stored at the terminal for the given time length, and the number of quay cranes that may be assigned to work simultaneously on the cargo of a ship. The water levels in port terminals, as well as over the various segments of river, canal, sea, and ocean routes are also significantly impacting the capacity of the associated physical-network arcs in terms of both the vessels that may be accommodated for berthing or navigation, and the weight of the cargo those vessels may carry (bridge heights may also limit these measures).

Several capacity measures are generally defined for an arc of the physical network, including of the same type when, for example, several categories of commodities or service types of the same mode (e.g., railway trains with different priorities and nominal speeds) may use the infrastructure simultaneously. Rules are generally defined in the latter case, governing the cohabitation of several types of commodities or services, yielding particular demand or service-related performance measures . Consider, to illustrate, the case of several train-service types (e.g., passenger, express freight, regular freight) operating on the same physical arc, each type with particular priorities when meeting trains of a different type. The average travel time of each type of service is then given by a multivaried function dependent on the total number of each service type operating during the same time length on the arc (e.g. Crainic et al., 1984).

Notice, finally, that the geographical, topological, and season characteristics of the regions concerned by the transportation system under study may have an impact on the arc characteristics, particularly in terms of capacity. Thus, for example, speed and maximum load capacity generally decrease in mountainous terrain (unless additional power units are assigned to services) or on sinuous physical paths. On the other hand, droughts, which are increasingly long and strong due to climate change, decrease the water levels of rivers and canals, jeopardizing plans and traffic.

Resources. We refer to resources required to operate "in the field", that is, manning services and terminal operations on services and freight. Notice that the term "assets" is often used in the industry and part of the literature. We prefer the term "resource" to refer to the human resources and material assets of the system. Also notice that, office resources are generally not part of the process and SND models, and are not addressed in this chapter.

Two main categories of physical resources are generally considered, *moving* and *terminal*. The former include vehicles (e.g., trucks, ships, barges, and airplanes), traction/power units (e.g., road tractors and locomotives), and hauling units (e.g., trailers, railcars/wagons, and containers). Terminal-handling resources may be more of less fixed, as the quay cranes in port terminals, or mobile, as the yard locomotives in railway networks.

The evolution of smart cities and urban-freight transport and distribution have seen the emergence and continuously-increasing deployment of several additional types of moving resources (e.g., Marcucci et al., 2024), which is reflected in the SND models targeting the planning of City Logistics systems (e.g., Marcucci et al., 2024; Crainic et al., 2023b,c). One finds among those resources electric and hydrogen-powered vans, railcars, as well as public-transport vehicles (e.g., buses, trolleybuses, tramways, subway cars) used partially or completely to transport freight. Operating the last kilometers, one also finds human-powered and electric cargo bikes, drones, and robots.

Human resources are also required to operate the system, terminals and services, in particular. We use the term *crew* when referring the group of people required to operate a service or taking care of specific terminal tasks. The composition of the crew varies with the mode, service and activity type, distance and time to operate. Its training and qualification determines the type of material resources and services the crew can handle.

Assigning human resources to services provides the means to execute them (except when automated vehicles are providing the complete service). On the other hand, assigning material resources to services define their characteristics. Power units determine how much can be hauled in terms of combined load and vehicle weight on each type of physical-network arc. Similarly, the loading/hauling units used determine the capacity of the service. Consequently, the resource management and resource-to-service assignment issues are very important for the definition, optimization, and performance of the carrier and its service network.

We address these issues in the context of tactical planning and SND modeling in Sections 3.3 and 5. We complete this subsection with a few observations:

1. The availability of each type of resource is limited; the term limited *fleet size* is generally used with respect to moving material resources. In general, the more costly the resource, the fewer the number of units available, e.g., locomotives are much fewer than railcars, planes and ships are fewer than trucks, tractors, and trailers.
2. It is noteworthy that, the transportation industry is facing an increasingly serious manpower issue. This impacts how drivers are asked to operate. To illustrate, consider the LTL relay-based networks (Crainic and Roy, 1992; Bakir et al., 2021). In a relay-network, drivers operate their vehicles for about half the work day until a meeting/relay point, where several services meet. Drivers then swap vehicles or trailers, each returning “home" to his/hers base terminal by the end of the day. Similar rules may be observed in the U.S. rail industry (Balakrishnan et al., 2016).
3. Particular management rules govern the utilization of each type of resource. With respect to human resources, there are the union-firm working agreements, together with national or international work rules. Thus, for example, the driver working rules in the European Union, Canada, and the United States are very similar but do present particularities that must be explicitly taken into account when planning (see, e.g., Albinski et al., 2020; Ammann et al., 2024). As for material resources, various levels and intensity of inspection and maintenance activities are prescribed for each type, e.g., rail engines (Ahuja et al., 2005b; Vaidyanathan et al., 2008a,b; Ortiz-Astorquiza et al., 2021).

3.2 Demand

As generally stated in the literature and recalled previously (Section 2.2), carrier tactical planning and the associated SND methodology, target the organization of the services the consolidation-based carrier supplies to answer the estimated/forecast regular demand for transportation of its shipper customers. Moreover, this demand for transportation means a request to move a specific quantity of a given type of freight between two locations serviced by the carrier, at, possibly, given time moments. A few modeling clarifications are in order.

First, we refer to “freight type" to recall that each demand is for a particular product, with specific physical and transportation characteristics and requirements, including weight, size, fragility, risk (e.g., dangerous goods), as well as packaging, handling, and product-vehicle adequacy rules. In many applications, e.g., most LTL and postal cases, one assumes that all demand types may be loaded together in the vehicles of the modes considered, and that each demand is defined by its weight or volume, or both, according to the most important vehicle and terminal capacity measure included in the study. This approach simplifies operations significantly in practice and is increasingly contemplated not only in the usual long-haul transport, but also in the urban context through the utilization of intelligent, modular containers in City Logistics (e.g. Crainic et al., 2023a,b). For illustrative purposes and to simplify the presentation, unless specifically stated, the general formulations discussed in this chapter assume all products in the problem setting may be loaded, and consolidated together, in the vehicles considered.

A somewhat similar approach is used when freight is loaded into containers or railcars at the shipper location, before the loading units are brought to the origin terminal. The demand is then expressed in terms of number of loading units (e.g., railcars or containers) of specific types and characteristics (e.g., 20- and 40-feet long containers, and gondolas for grain, flat cars for containers or trailers, and boxcars for general cargo on rail). Specific product-to-vehicle assignment restrictions, e.g., fresh or frozen food requires special refrigerated vehicles, even though one could load them in boxcars or general containers, require specific modeling of the individual products and of the product-to-vehicle assignments (see, e.g., Crainic et al., 2009, 2021c, in the City Logistics planning context). Each OD shipper demand is then expressed as a vector of product quantities, together with (1,0)-valued product-to-vehicle assignment coefficients (i.e., = 1 if the product may be loaded into the vehicle type, 0, otherwise). Such coefficients may be general for the problem or particular for each shipper. It is noteworthy, however, that loading concerns are not included in most tactical-planning models. Only very recently, one notices interest in refining the demand-to-vehicle loading rules and constraints of SND models for both intermodal transportation (e.g. Morganti et al., 2020; Kienzle et al., 2024) and general consolidation-based transport (Bruni et al., 2024).

Demand is also characterized by economic and service elements. The latter generally involves the time allocated to the delivery of freight to destination, as well as, increasingly, when demand is to be available for transportation (Section 4). The former takes the form of the fee, *tariff*, shippers pay for the transportation of their goods, which is conditioned by a combination of freight type, distance, service requirements, and commercial understanding (e.g., long-term contracts offering discounts on regularity and volume). Penalties, for late deliveries and damage, for example, may be part of the commercial deal. The term *multi-commodity demand* is generally used to represent this diversity in the literature (Crainic et al., 2021a) and in this paper.

Second, the customer classification. Most contributions in the literature focus exclusively on the *regular* demand, as introduced in Section 2.2, which reflects general planning processes in industry. This is not, however, the only demand the carrier has the opportunity to service in actual operations. One may identify two other main customer categories. First, the so-called *potential* (or intermittent) demand corresponding to shippers without formal contracts or long-term understandings, but which are “known" to the carrier and which, from time ot time, call on the carrier’s services and which could be prompted for loads if need be (e.g., to fill up a vehicle which would otherwise return home empty). So-called *spot* customers make up the last category, including demand which, historically, pops up more or less unexpectedly during operations. While generally not accounted for in tactical-planning models, such categories gain in importance when exploring the market potential of our services (e.g. Andersen and Christiansen, 2009b), or in tactical planning of carriers and intermediary firms when customer selection is involved, particularly in the presence of revenue management policies, tariffs, and customer stratification definitions that have to be accounted for in the corresponding SND models (e.g. Bilegan et al., 2015, 2022; Taherkhani et al., 2022). Moreover, identifying sources for additional demand within the potential and spot customers plays an important role when explicitly addressing demand uncertainty.

Aggregation is also a core elements in defining the demand for tactical planning. Aggregation takes place at two levels simultaneously, geographic and shipper identification. With respect to the former, customer locations are assigned to a regional terminal or hub, their associated oputgoing and incoming demands being aggregated accordingly. Turning to shipper identification, one notices that customers are not individualized at the tactical planning level, except for very important ones with respect to freight volumes, revenues, personal relations, etc. Hence, the demands of shippers with similar characteristics in terms of type, economic, service, origin and destination are aggregated yielding the OD demand used in the SND formulation.

A final remark with respect to demand modeling. Trade is unbalanced among countries and regions and, consequently, so is the demand for particular vehicle types. Hence, even though moving empty vehicles is costly and carriers aim to minimize them, one needs to balance the vehicle flows, and these decisions have to be echoed in the tactical-planning methods. Particularly when neglecting them yields a significant underestimation of the vehicle, or convoy (e.g., container intermodal transport by ship and train) oreven infrastructure (e.g., ports and rail lines) utilization. This is increasingly performed by including resource-management concerns into SND models (Section 5). When this is not the case, one can still account for the empty flows by estimating origin-destination “empty-vehicle" volumes and include them in the demand definition (e.g., Crainic et al., 1984; Powell, 1986a)

3.3 Service Network

Services are defined on the physical network, given the available resources and according to the carrier’s operations policies and customer service-quality targets. Main ingredients in defining services are the physical route, resources required and capacity, evaluation measures, and operation type.

Service route. As indicated previously, a service $\sigma \in \Sigma$ follows a path in the physical network from its origin $O(\sigma)$ to its destination terminal $O(\sigma)$. Several other terminals may be located along this path. A so-called *direct* service would generally pass these terminals without stopping. The service route is then represented as a single arc $a \in \mathscr{A}$ of the potential service network of a static SND formulation, as defined in Section 2.4.

A multi-leg service halts at intermediary terminals on its route to load and unload cargo into/from its vehicle (not only at the origin and destination terminals of the commodity, but also for consolidation and transfer purposes) or, when convoys are involved (e.g., rail, road, and barge trains), to pick up or drop off individual or groups of vehicles, e.g., car and blocks for railroads and trailers for LTL motor carriers operating multi-trailer road trains. The service route is then described by the sequences of $n(\sigma)$ stops, origin, intermediary, and destination, and service legs connecting them. The single-leg, direct, service case has $n(\sigma) = 2$. Let $\mathscr{N}^{\text{PH}}(\sigma) = \{\eta_i(\sigma) \mid i = 0, \ldots, n(\sigma), O(\sigma) = \eta_0, D(\sigma) = \eta_{n(\sigma)}\}$ be the stop sequence of service $\sigma \in \Sigma$. Then, the *service leg* $l_i(\sigma) = (\eta_{i-1}, \eta_i)$ is defined as the sub-path connecting the consecutive terminals η_{i-1}, $\eta_i \in \mathscr{N}^{\text{PH}}(\sigma)$ of the route of service σ, with $\mathscr{L}(\sigma) = \{l_i(\sigma), i = 1, \ldots, n(\sigma)\}, \sigma \in \Sigma$.

Multi-leg services yield several arcs in the potential service network G, each service leg corresponding to an arc of the service network, that is, $\mathscr{A} = \mathscr{L} = \bigcup_{\sigma \in \Sigma} \mathscr{L}(\sigma)$. Let $a(l_i(\sigma))$ and $l_i(\sigma)(a)$ stand for the link defined by leg $i \in \mathscr{L}(\sigma)$ of service $\sigma \in \Sigma$ and the service leg defining arc $a \in \mathscr{A}$, respectively. Then $u_a = u_{l_i(\sigma)(a)}$ and $u_a^k = u^k_{l_i(\sigma)(a)}$ in constraints (3) - (4) of the SND formulations.

It is noteworthy that other operations besides those mentioned above may be carried out at intermediary stop. Such operations include compulsory inspection of vehicles and cargo, switching of equipment (for example, at borders or when the infrastructure requires it, e.g., electric tension or rail gauge for rail transport, Andersen et al., 2009b; Andersen and Christiansen, 2009a), and crew changes, which may be particularly observed in the LTL motor-carrier and rail industries. Indeed, most countries, as well as the European Community, impose strict truck-driving work rules. Those restrict the length of continuously-driving periods, as well as the cumulative driving time over certain time periods (one, two, three days, for example), imposing rest periods proportional to these cumulative values. Implementing the relay-based operations mentioned previously mentioned is part of the answer to this issue.

Those must then undertake appropriately long rest periods, which used to be common practice but are increasingly less so, because of costs and the requirements of drivers (who are in short supply). To address this issue, the service (truck) stops at an intermediate location, wherein an exchange of drivers occurs, each returning to his/;hers base terminal with the other driver's "vehicle". Although not new (Crainic and Roy, 1992), the utilization of relay-based service networks is increasingly steadily, but without a similar research-effort intensity. Similar situations are encountered in melting snow and unfreezing of the ground outside will rail transportation. In the U.S., for example, a railroad territory is divided into so-called districts, and trains circulating through a district must be operated by crews from the same district, following complex regulatory and union rules (Balakrishnan et al., 2016). Hence, crew changes must take place at appropriately selected stops. Note that, in all cases, such intermediate locations need not be a terminal on the service route in the physical network, but any suitable site. Thus, e.g., driver exchanges may occur at rest stops on a highway.

Resources and capacity. Associated with service σ is a global capacity u_σ, which can be commodity specific, u^k_σ, as well as leg specific $u_{l_i(\sigma)}$ ($u^k_{l_i(\sigma)}$), $l_i(\sigma) \in \mathscr{L}(\sigma)$, representing the total volume of freight the service may load and haul on the leg. The latter are generally defined when there are significant differences in the infrastructure state, e.g., the maximum permitted weight of loaded motor vehicles is reduced on roads and highways in northern parts of the globe when Spring thaw melts snow and unfreezes the ground (it is noteworthy that climate change brings rapid negative consequences to the northern regions through unfreezing of the permafrost supporting all settlements and infrastructures), or the power required to haul the assigned loads, e.g., when climbing or descending mountainous terrain. Clearly, $u_\sigma = \min\{u_{l_i(\sigma)}, l_i(\sigma) \in \mathscr{L}(\sigma)\}$ in such cases (same note for commodity-specific capacities).

The service capacity is determined by the resources allocated to it. First, by the actual loading capacity of the vehicle, or vehicles, associated to the service, e.g., the vessel, truck, or trailer size, as well as the number of such vehicles when rail, truck, or barge trains are part of the problem setting. Second, by the power and number of traction units assigned to the service, e.g., engines for railroads and tractors for LTL motor carriers. Consider, to illustrate, that the capacity of North American trains operating on the main lines has been significantly increased by adding a locomotive, sometime located in the middle of the convoy. Given the high acquisition and operation/maintenance/depreciation costs of such material resources, it is increasingly important to address resource-to-service and resource management concerns at the tactical-planning level.

The limits on available human resources need to be accounted for as well, as they impact not only the costs, but also the type and number of services a carrier may deploy. Consider, for example, the scarcity of truck drivers, particularly for the long-haul movements requiring several days out of home. The particular qualifications required to conduct vehicles and convoys for most modes, and the strict work rules controlling crew schedules in all transportation modes are also constraining the design of the service network. We discuss the integration of resource-management concerns into SND models in Section 5.

Service type. Different types of services may be operated on the same physical route, mainly differentiated by the service duration (and, when relevant, priority) and related service-quality targets in terms of predictability/regularity, i.e., achieving the planned duration and, hence, the promised delivery dates to shippers, at a given level over the planning horizon (percentage of number of on-time travel and delivery with respect to the total number performed).

The service duration is the total time required to reach the destination terminal once the service leaves its origin terminal, and is the sum of the travel times of the service legs and the stops at intermediary terminals (time at stops for equipment or crew monitoring or change are normally included in the leg travel times). The leg travel time is primarily determined by the carrier's performance measures and goals. The carrier may, for example, offer both “express" and “regular" services between two cities to better capture the different types of shipper requirements. Such decisions are generally part of revenue management strategies (e.g., Bilegan et al., 2015, 2022). Of course, “express" means shorter duration, with possibly shorter stops at terminals, involving higher costs, more powerful traction and energy consumption (and, possibly, higher emissions), yielding higher tariffs. While deciding on implementing such service differentiation is a strategic decision and its implementation heavily involves marketing and sales, it also has impacts on planning through the need to assign and manage the appropriate resources and to account for the associated performance measures in the planning processes and models.

Travel times, and physical routes, may also be influenced by environmental and energy-saving concerns. There is not, as yet, much work done in this area. Currently, most efforts target the introduction of more environment-friendly transportation modes and infrastructure, such as the cargo bikes, electric and hydrogen vans, automated vehicles, and charging stations one finds in City Logistics systems. An interesting emerging research field addresses the case when motor vehicles are grouped in platoons for part of their journeys, with promises to save on energy and environmental negative impacts, as well as on crew costs (e.g., Scherr et al., 2019; Albinski et al., 2020; Scherr et al., 2022; Ammann et al., 2024). Bauer et al. (2010) present one of the few studies incorporating environment-related costs into freight transportation planning. They discuss quite at length the difficulty of estimating emissions and impacts in general and the additional challenge of building measures which may be included into a SND formulation. Experimentation on an intermodal rail case (from Andersen et al., 2009b) point to the trade offs between operation and environmental costs and the change in service network when the latter are given a significant economic position.

The service type, particularly in terms of predictability and customer delivery-date satisfaction, is also dependent on how the service is dispatched. Not long ago, many services were dispatched out of their origin (or, even, intermediary) terminals “when full", the schedule, when existing, being indicative only. Such an approach aims to maximize the utilization of the vehicle or convoy and minimize operation costs, with very loose travel time targets. Operating according to more stringent schedules offers the means to aim for higher service-quality targets.

Frequency-based operation implements a somewhat flexible schedule through a combination of “leave when full as much as possible" and “dispatch another vehicle on the same service, if needed" policies and the definition of definite times intervals within which this combination is applied. Thus, the “same service" may be operated several times during a “normal" activity period, e.g., several trucks may be dispatched during the day between the same two cities, or several ships (or the same ship performing back-and-forth services) may sail during the week or the month. LTL motor carriers operate according to such a policy by defining shorter time intervals over a normal activity day, each interval with a particular frequency (e.g., between the same two cities, 2 departures in the morning, 3 between noon and 15:00, and 10 between 15:00 and 19:00), reflecting the operation practice of receiving loads in the morning and aiming for outgoing vehicles leave on time to get to their next terminal early “the next day", to facilitate consolidation, transfer, and delivery activities.

Precise scheduling is more the norm when the infrastructure is shared with other modes or carriers, e.g., intermodal shipping, air, and rail transportation. In such cases, each service is characterized by a schedule indicating the arrival and departure times at each terminal defining its route. These times may be precise or within a specific interval. To simplify the presentation, but without loss of generality, we assume the former in the rest of this chapter.

Then, in such *time-dependent* problem settings, the OD demand $k \in \mathscr{K}$ is characterized by its *availability time* $\alpha(k)$ at origin $O(k)$ and due date $\beta(k)$ at destination $D(k)$. The schedule characterizing service $\sigma \in \Sigma$ gives the departure and arrival times, $\alpha(\eta_i)$ and $\beta(\eta_i)$, at the origin and destination terminals, respectively, of each of its legs $l_i(\sigma) \in \mathscr{L}(\sigma)$. The service total duration is noted $\tau(\sigma)$, and includes the time spent in terminals and the moving time associated to each leg $\tau(l_i(\sigma))$.

Addressing scheduled services, be it according to frequencies or to more precise arrival & departing times, constitute a significant extension of the SND methodology, yielding time-dependent, or time-sensitive, *Scheduled Service Network Design*, *SSND*, models discussed in Section 4. It is also reflected in the evaluation and performance measures considered in the problem setting and the corresponding formulations.

Evaluation measures. Following general planning and decision-making in transportation, *cost* is the most encountered evaluation measure in tactical planning and SND methodology. It is a generic term used to capture the monetary / economic measures associated to operating the system and executing the tactical plan and, thus, to evaluate alternatives and make choices. *Time* or *delay* measures are also found, and increasingly so, reflecting the shipper service-quality requirements. We focus on these two categories in this chapter.

As already indicated (Section 2.4), the SND objective function is generally composed of two cost terms, a so-called fixed *service-network design* cost and a *service-network utilization* cost incurred to transport the given demand. In network-design vocabulary, the “fixed" qualification of the former aims to indicated that those costs are to be paid if selected, independently of the flow assigned to corresponding arcs. In the vast majority of SND applications and contributions, both terms are linear, as illustrated in functions (1) and (20), i.e., both terms are computed multiplying the appropriated decision variable, service design and commodity flow, respectively, with the associated unit cost, and summing up the result over all services and network arcs, respectively. In most cases, these unit costs are given constants, $f_\sigma, \sigma \in \Sigma$, and $c^k_a, k \in \mathscr{K}, a \in \mathscr{A}$, respectively.

The service fixed cost capture the economics of operating each occurrence of the service. It generally accounts both for the office cost to set up the service (e.g., management, marketing, etc.) and the in-the-field cost of operations. The latter depends upon the application. It may involve, for example, the acquisition-depreciation-maintenance cost of the power units or vehicles, particularly when resources and time are not explicitly included in the formulation. Energy costs to haul the nominal service load over the route distance at the planned speed, are generally included, as are the crew-related costs.

The unit arc commodity cost reflects the part of the total estimated service-operating cost that relates to each particular commodity. Such a cost accounts for the length of the arc, and may reflect, e.g., the weight of each unit of product making up the commodity, the particular requirements in terms of vehicle used or handling procedures. Similarly to the fixed-cost case, the unit flow cost may capture the potential environmental impact of risk (for dangerous goods, for example) of particular commodities. It may also capture the impact of the travel and stop-at-terminal times on the particular commodity according to its priority (the higher the priority or the importance of rapid delivery, the higher the cost).

The trip duration, the respect of delivery-date engagements, and, more generally, the delays incurred by vehicles, convoys, and freight, due to congestion and operational policies in terminals and on the road, are generally used as a measure of service quality and increasingly reflected in the SND “costs" and objective functions.

In its basic form, and the one most widely encountered in the literature, stopping and working times in terminals, as well as travel times between terminals are assumed known and fixed. Such measures are then usually translated in monetary values. The resulting time-related service and unit-commodity costs may be then either added to the operational ones, as indicated above, or kept as separate terms of the objective function. Choosing one or the other approach does not impact the optimization model and results. The second one may, however, provide more flexible approach to managerial studies when one desires to study, e.g., the impact of modifying the relative value of time and delays relative to direct cost of operations.

Such linear formulations do not always reflect reality adequately, however. Consider the case, for example, when vehicles or convoys meet on restricted infrastructure such as rail lines or canals, and one must “step aside" (on a side track in the rail case) to let the other one pass by or overtake. A particular delay is then experienced by each participating vehicle or service, according to the priority of its service type and the traffic intensity, that is, with the number and service types of the other vehicles or services involved. The amplitude of the delay generally increases non-linearly with the traffic intensity. Similar situations may be observed with increasing volumes of freight or loading units trying to use the same capacitated infrastructure such as classification rail yards and maritime ports.

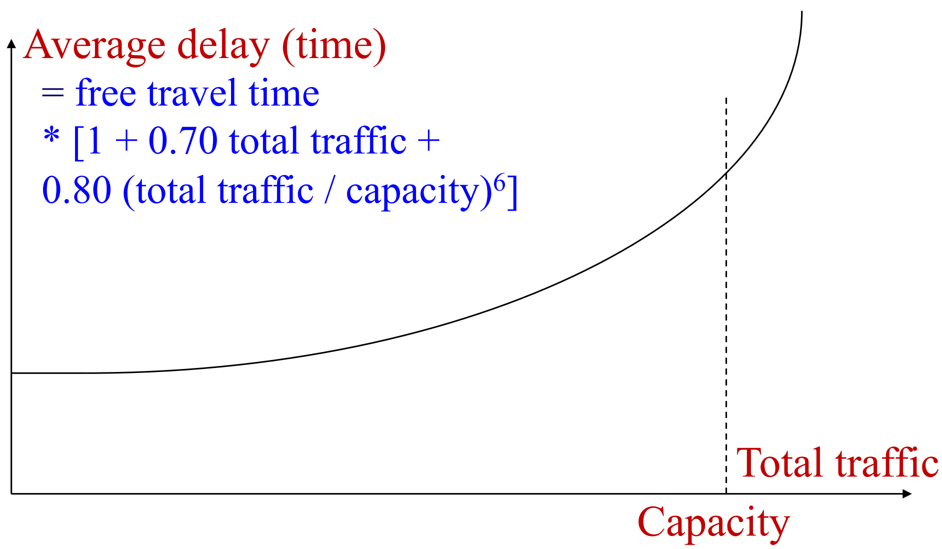

Figure 3: Average delay under congestion conditions

We refer to such cases under the generic name of congestion conditions and to the related time representation as *average* or *unit delay*. Congestion functions are widely used in road traffic modeling and analyses. They approximate traffic conditions and impact on travel time for use in optimization and simulation methods. Hence, they generally take the form of continuous convex function representing the expected average travel time over a given road segment, given the total traffic and a certain measure of the capacity in number of vehicle using the road for a certain time duration. Similar approaches are used for estimating delays in congested facilities use in for tactical and strategic freight transportation planning (e.g., Crainic et al., 1984, 1990b).

Starting generally from queuing models (engineering procedures is some cases), continuous and convex functions are approximated to represent the behaviour of the facility under study (e.g., Crainic and Gendreau, 1986; Powell, 1986b; Powell and Humblet, 1986; Crainic and Quérin, 1988). Figure 3 illustrates such an idealized congestion function. The traffic may be in terms of vehicles, convoys, or freight, e.g., railcars in classification terminals and ships in ports, train services on rail tracks, and load classification and consolidation in LTL terminals, respectively. Notice that, the “capacity" measure used such function is rarely the actual physical capacity of the corresponding facility. It is rather an “ideal" limit, in terms of managerial acceptance of delays and costs, of the maximum traffic the facility should handle while operating under normal conditions. One identifies three main parts to a volume-delay function. In the first part, the function reflects the fact that there is a non-compressible duration required to perform an activity and that this duration is fairly constant when the traffic volume is low. The delay than gradually increases with traffic, more or less linearly in the second part, and then non-linearly in the third. This last part, where the function increases more or less sharply with traffic, plays several roles, where it 1) reflects the fact that the average delays increase non-linearly with traffic and models the actual behavior of the system before the ideal capacity; 2) together with the definition of the ideal capacity, captures the managerial objective of diverting traffic to other paths before the system becomes too congested (i.e., the traffic is larger than the ideal capacity); 3) models the planning reality that, the capacity of a vehicle or facility may be exceeded at a cost, increasingly higher as the overflow grows, representing the recourse to additional capacity, rented or own; 4) finally, through the sharpness of the curve near and passed the capacity, guides the optimization model to direct flows to other paths and facilities to avoid congestion.

These ideas may be refined to guide the representation of many restrictions and conditions in a more realistic way. The fundamental concept is that, while limits are hard conditions in actual operations, one does not necessarily need to model this hardness straightforwardly for an optimization formulation. One does not desire the method to crash because a capacity is attained or exceeded. One would rather prefer a model that searches for more balanced alternatives, in which vehicle or freight flows are directed toward other facilities or services. The same situation concerns the delivery dues dates of shipments, for which, in most cases, one may deliver late paying a more or less high cost.

Penalty terms are the usual instrument to model such situations. Formula (23) illustrates this approach for the service penalty, for the basic SND formulation (1) - (6). Let $p_{\sigma(a)}$ be the unit penalty cost for exceeding the capacity of service $\sigma \in \Sigma$. Then the term in (23) stands for the total penalty cost of the system and is to be added to the objective function (1), while the linking-capacity constraints (3) are to be dropped.

$$\sum_{a\in\mathscr{A}} p_{\sigma(a)} \left(\min \left\{ 0, u_a y_{\sigma(a)} - \sum_{k\in\mathscr{K}} x_a^k \right\} \right)^2 \tag{23}$$

Trade offs between the cost of increasing the level of service and the extra costs of insufficient capacity may then be addressed while the associated mathematical programming problem is solved.

We complete this section by noticing that averages often do not tell the full story. This may be the case with, for example, transportation delays. Often, the goal is not only rapid delivery but also consistent, reliable service. The variance of the total service or itinerary duration may then be used to penalize unreliable operations (Crainic, 2003). Equation (24) illustrates this approach for the case when service-quality targets are announced and a path-based formulation (7) - (12) is used. For each commodity $k \in \mathscr{K}$, let H_k be the target delivery objective (e.g., 24 hours), p^k the penalty for not complying with this service objective, and n_k the reliability requirements (e.g., target must be achieved for 90% of deliveries) the carrier defines for the commodity (market). The term (24) then computes the total penalty the carrier contemplates when the expected itinerary duration, adjusted for its standard deviation, does not comply with the service objective. (Notice that, (24) assumes that arc travel times are independent, which is verified in deterministic version; correlations must be accounted for, when independence is not verified.)

$$\sum_{k\in\mathscr{K}} p^k \sum_{\pi\in\Pi^k} \left(\min \left\{ 0, H_k - E_\pi^k(y,h) - n_k SD_\pi^k(y,h) \right\} \right)^2 \tag{24}$$

4 Modeling Time

The basic service network design formulations discussed in the previous sections do not integrate time-related concerns, at least not explicitly. They do not ignore it, of course.

Recall that, the length of the tactical planning horizon is determined by the homogeneity of that time duration, in terms of regular demand and activities, in a stable environment. The season length varies with the carrier type, but may also vary with the climate, e.g., between dry and rainy, cold (winter) and warn (no-winter) seasons. Recall also that, the schedule length, that is, the time interval covered by the tactical plan is much shorter than the season length, being largely determined by the repetition pattern of the regular demand. The activities planned for this schedule length are then repeatedly executed for the duration of the tactical planning horizon.

The basic formulations address situations where there is no variation in demand for the duration of the schedule length. Either the schedule length is short and one assumes that everything “happens simultaneously", or the demand arrivals and service departures are assumed to be uniformly distributed over the schedule length for longer time spans (e.g., the 1440 minutes of the week, Crainic et al., 1984). Because the time attributes of the various problem elements are not explicitly included, those SND models are qualified as *static*.

Notice that, even though identified as “static", the problem settings and models involving service frequencies (formulated as $y_\sigma \in \mathbb{Z}_+, \sigma \in \Sigma$ in constraints (21)) are modelling time and activities in time, assuming that the multiple executions/departures of the service are equally spread out over the schedule length. See, for example, Crainic and Rousseau (1986); Crainic et al. (1984); Roy and Delorme (1989), where service frequencies are modeled as decisions, and Powell and Sheffi (1983) where service frequencies are extracted out of the optimization results.

Tactical planning problem settings and SND formulation are increasingly and consistently addressing the variations of demand in time and the scheduling of the services selected to answer the transportation requests. The problem settings and formulations explicitly identifying time-related attributes of the system components and operations are qualified as *time-dependent* (the term *time-sensitive* is also used). The most frequent time-related attributes of demand are the *availability time* $\alpha(k)$ at the origin $O(k)$ of $k \in \mathscr{K}$, and the *due date* $\beta(k)$ at destination $D(k)$. The latter is sometimes refined as an interval around an ideal delivery date, the lower and upper limits of the interval modeling the earliest and latest time the consignee is ready to accept delivery; penalty costs may be associated to these limits, late deliveries, in particular.

Similarly, services are characterized by a *schedule* indicating the departure and arrival times, $\alpha(\eta_i)$ and $\beta(\eta_i), i = 1, \ldots, n(\sigma)$, respectively, at each of the terminals $\mathscr{N}^{\mathrm{PH}}(\sigma)$ on their routes. Services are further characterized by a total duration $\tau(\sigma)$, that includes the time spent in terminals and the moving time associated to each leg $\tau(l_i(\sigma))$. Schedules may be strict, as for most European and Canadian railroads, somewhat flexible (e.g.,)by specifying a day and time interval), or more of an “indicative" nature, the schedule being eventually modified to account for how much freight is already loaded.

How does one represent “time" in the context of SND modelling for tactical planning of consolidation-based carriers? In other words, how does one represent *within the schedule* *length* the time-attributes of the system elements and the associated events, e.g., demand becoming available at origin, the starting and duration of a classification activity in a terminal, and the scheduled departure of a service?

To answer such questions and capture the time-related characteristics of demand and service, SND models are generally defined on a *time-space network* $\mathscr{G} = (\mathscr{N}, \mathscr{A})$, which is typically built by extending the network $\mathscr{G}^{\mathrm{PH}} = (\mathscr{N}^{\mathrm{PH}}, \mathscr{A}^{\mathrm{PH}})$ along the dimension of time for the duration of the schedule length. The service legs provide the (potential) arcs supporting the movements through space and time of the vehicles and convoys of the various modes considered, while itineraries perform the same role for the transportation of time-dependent demand. When formulated on such a network, the SND is often referred to as a *Scheduled Service Network Design* (*SSND*) model.

A time-space network is often built by partitioning the schedule length into non-overlapping periods of time. Then, as it is standard modeling practice in multi-period optimization, all activities taking place at terminals during a period are assumed to occur at the same time, be it the beginning (most often) or end of the period. This approach is also known under the term *time discretization*. Let T be the schedule length. Let $\{0, 1, \ldots, |\mathscr{T}| - 1, |\mathscr{T}| = T\}$ be the set of time instants t that partition the schedule length into $|\mathscr{T}|$ periods, and $\mathscr{T}$ the ordered set of those periods. Period t is then defined as $(t-1, t)$, all events taking place during this period being assigned to time instant t.

The granularity of the partition and the definition of each period are normally governed by the characteristics of the system elements and operation practice. Most applications in the literature, however, implement the classical approach, first introduced by Ford and Fulkerson (1958), according to which all periods are of the same length and apply to all the nodes of the network. The node set of the time-space network, are defined in such cases as $\mathscr{N} = \{\eta_t, \eta \in \mathscr{N}^{\mathrm{PH}},\ t = 0, \ldots, |\mathscr{T}|\}$, including copies of all the terminals in the physical network at all the time periods (instants) defined. Figure 4 illustrates this modeling approach for a tiny network of four terminals and a schedule length with $T = 4$ partitioned into four equal-length periods. Each of the four services is identified by a color and line pattern on the physical network at the left of the figure. The service schedules are displayed, by service leg, next. The corresponding deployment in time and space of each service, according to those schedules, is displayed on the time-space diagram at the right of the figure.3

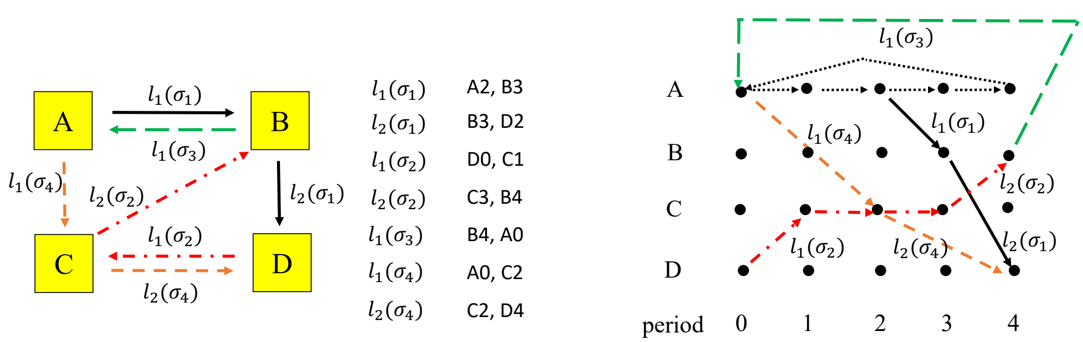

Figure 4: Time-space network representation

The arc set $\mathscr{A}$ consists of two types of arcs, *moving* and *holding*. Moving arcs correspond to the service legs according to their schedules. Specifically, the moving arc $a \in \mathscr{A}$, standing for service leg $l_i(\sigma) = (\eta_{i-1}, \eta_i), i = 0, \ldots, n(\sigma), \sigma \in \Sigma$, is defined as $a = (\eta_{\alpha(\eta_{(i-1)})}, \eta_{\beta(\eta_i)})$, representing the departure of the service leg from its origin terminal η_{i-1} at time instant $\alpha(\eta_{(i-1)})$ and arriving at its destination terminal η_i at time $\beta(\eta_i)$. Holding arcs represent the possibility to wait, to "hold" for one period goods or resources at terminals. Such an arc is thus of the form $a = (\eta_t, \eta_{n+1}), \eta \in \mathscr{N}^{\mathrm{PH}}$, $t = 0, \ldots, |\mathscr{T}|$. There are no fixed costs associated with holding arcs. A number of capacity and unit cost parameters may be defined, however, to represent the handling or warehousing capabilities of the terminal and costs of this handling or of keeping resources idling. Figure 4 illustrates seven moving arcs for the seven service legs, two holding arcs at terminal C representing the length of the stop service σ_2 performs at terminal C, and the five holding arcs linking successive representations in time of terminal A (the other holding arcs at terminals B, D, and D are not shown for clarity sake).

Figure 4 includes two arcs, one moving from terminal B at time 4 to terminal A at time 0, and the holding arcs from time 4 to time 0 at terminal A, which seem to go backward in time. The apparent time travel is only an illusion, however, produced by the graphic display of a modeling device to account for, on the one hand, the repetitiveness of the tactical plan answering the regularity of demand over the season and, on the other hand, the fact that not all demands and services have their initial and terminal instants within the schedule length; some may start during the previous application of the plan and terminate currently; others, start during the current application of the plan, but terminate during the following one. These issues are addressed by having the moving and holding arcs modeling activities that would end during the next application of the plan *wrap-around*. To avoid the apparent time-travel paradox, the time computations are performed $modulo(T)$ (see, e.g., Crainic and Hewitt, 2021, for details).

The SSND formulations then take the form of the corresponding basic SND models (Section 2.4) applied to a time-space network $\mathscr{G}$ as defined above. The SSND considers the same two sets of decision variables, selecting scheduled services (with frequencies, in some cases, representing the number of simultaneous departures of the service) and building itineraries in the service network for demand-flow distribution. The SSND constraint sets are analog to those of the SND, but they are written to account for the time attributes of the system, e.g., for the flow-balance equations defined according to the availability instant and the delivery due date.

The differences and challenges come from applications to particular problem settings and the dimensions of those problems and of the corresponding formulations. Three illustrations follow (the interested reader is directe.d to Crainic and Rei, 2024, as well as the surveys indicated in Section 2):

1. Zhu et al. (2014) propose a multi-layer time-space SSND model, which appears to be the first comprehensive rail SSND formulation, which selects the scheduled service network, the car classification policies, the blocks to build in each terminal, with their routes within the service network, the train makeup, and the demand itineraries using these services and blocks. The authors also introduce a matheuristic solution methodology combining slope scaling, a dynamic block-generation mechanism, long-term memory-based perturbation strategies, and an ellipsoidal search, i.e., a new intensification mechanism to

thoroughly explore very large neighborhoods of elite solutions in an efficient way using information from the history of the search.

2. Jarrah et al. (2009) propose a time-space network SSND formulation for the LTL service-design and flow planning problem with explicit service commitments (delivery days). Based on LTL-carrier no-split (single itinerary) shipment policies (Powell and Koskosidis, 1992), the authors formulate the problem using the in-tree structure of the problem, and propose a meta-heuristic integrating a dynamic in-tree generation procedure (see, e.g., Erera et al., 2013; Lindsey et al., 2016, for algorithmic developments).
3. Crainic et al. (2009) introduce the first modeling framework for the short to medium-term planning of two-tier City Logistics systems, which takes the form of a path-based formulation combining a multi-product, multi-modal SSND model for the first tier, and a scheduled, multi-depot, multi-tour, heterogeneous vehicle routing with time windows formulation for the second tier. Demand itineraries link the two tiers and provide the synchronization environment (see, e.g., Crainic et al., 2016; Fontaine et al., 2021, for extentions to uncertainty modeling and algorithmic developments).

The representation of time and the time-space formulations raise a number of issues and challenges and, while quite a volume of work has been already dedicated to these issues, this is still a very broad and rich research area.

One notices, for example, that the schedule-length partition does not need to be the same at all terminals. Indeed, the relative importance of the terminal with respect to the overall work load may point to the need to have a fine granularity for high-utilization terminals, including the hubs. Many services start, stop, or terminate, at such terminals, over most of the schedule length duration, and many demand flows need to be handled. In contrast, several smaller regional terminals may have to operate at less intensive levels and at certain time moments only. Figure 5 illustrates such a representation. It displays a number of regional terminals (e.g., numbers 3, 5, and 16) where very few feeder-like services originate and terminate to connect them to hubs. It also displays the more intense time intervals hubs (e.g., terminals 2, 6, 8, and 15) operate, together with the idle periods in-between those intensive-work periods. The figure also displays the moving arcs of the services connecting the terminals in time and space. The drawing is inspired by the case of freight-rail planning in Europe where freight trains must be scheduled within particular time intervals (so-called time canals) between periods of intense passenger traffic (Pedersen and Crainic, 2007). Services hence arrive at a terminal the beginning of a work period and leave at the end. In more general setting. this restriction may be relaxed.

Mini-time-space networks may then be used to appropriately model each terminal in time, these mini-networks being then connected through the services arriving and departing at the node during the interval. Figure 6 illustrates such a node representation. It shows three services arriving at the terminal, two of which are at their destination (inbound full red arrows), the third being at an intermediary stop (dashed red arrows). Two outbound services are displayed, one at its origin (outbound full red arrow), the other departing after its scheduled stop. The figure also displays arcs modeling particular terminal activities including, 1) cargo (and blocks for a railway yard) transfer between inbound and outbound services (dotted blue arrows); 2) the possibility for freight on the stopping service to stay on board (the dashed red arrow); 3) cargo

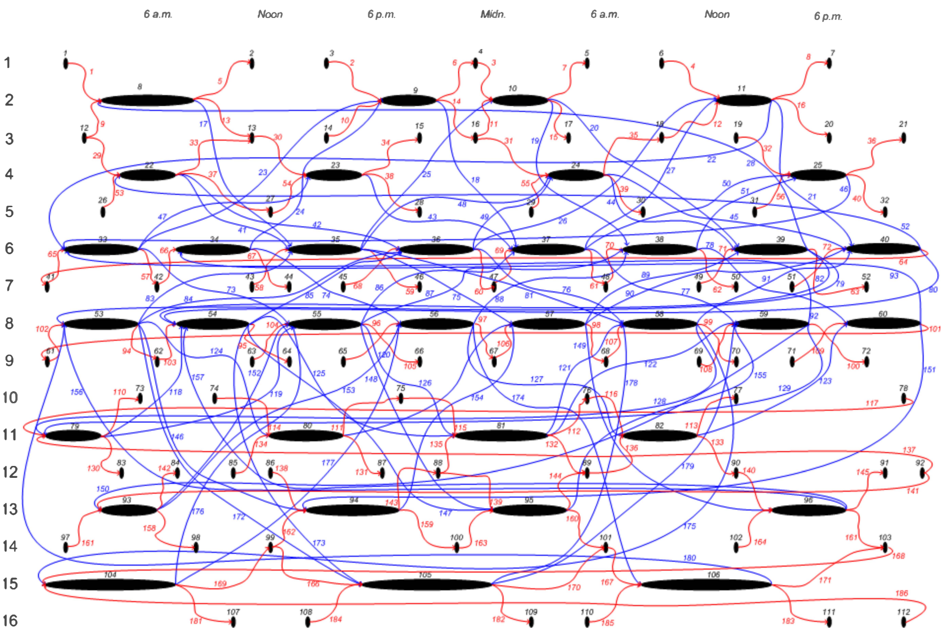

Figure 5: Time-space network with non uniform discretization

at destination being unloaded for delivery (outbound dotted black arrow); 4) the possibility to unload cargo (or dismantle blocks at destination) for classification and consolidation prior to being put on a departing service (full black arrows); 5) freight at origin becoming available and either going through the classification and consolidation operation, or being put on hold until this operation is to be performed (dotted black arrows). Notice that a continuous-time representation is implied for the activities modeled by the arcs of a mini-network, the particular duration of each being at attribute of the arc. A discreet-time representation proper to the particular terminal and time interval considered is also possible.

One may also observe that, the partitioning of the schedule-length is not required when the schedules of potential services are strict. One may rather use the arrival and departure times of services at terminals to create the time instants of the time-space network. Then, the physical nodes are duplicated at relevant time instants only (that is, when the event takes place at the terminal), and the availability time of each demand at its origin terminal is “projected" on the first time instant following arrival (Morganti et al., 2020; Kienzle et al., 2024). One still models on a time-space network in such a representation, but determining an appropriate granularity of partition in not a real issue.

Different approaches may be called upon when service schedules are not so strict (or, when time-related uncertainty is considered) and one is interested in the number of departures within a given time interval, to ensure timely delivery of demands at destinations, rather than in the precise departure instant, which may vary when the plan is executed repeatedly. Modelling the schedule length and associated events using so-called continuous time is such an approach. The timing of events related to service, and itinerary, operations then become part of the decision variables, and a significant amount of constraints governing arrivals, departures, and activity

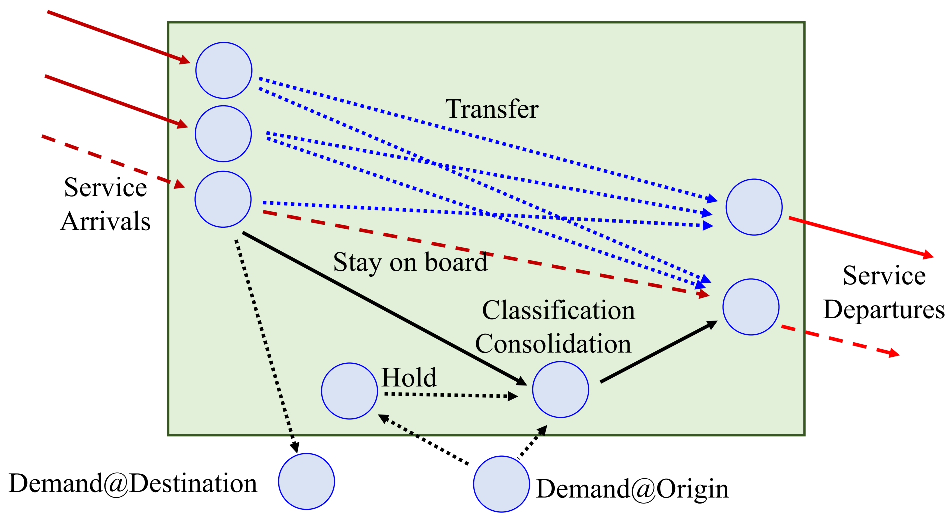

Figure 6: Time-space network with non uniform discretization

synchronization at terminals has to be included (e.g., Lange et al., 2024). More research is required in this area.

A coarse-granularity-based approach may be used for short schedule lengths (e.g., a working day as encountered in many LTL applications). The periods, generally of unequal length, may then be tailored to fit the arrival pattern of demand and the desired pattern of departures to ensure on-time arrival at the next stops of outbound services (e.g., three periods, morning, early afternoon, late afternoon / early evening). In such a case, the schedule of the service would indicate the coarse time of the day when it is supposed to leave or arrive at the terminal, and a frequency-based formulation associated to each period would determine the number of service departures.

In all the other many cases where time discretization is the preferred methodology for SSND, the issue of the partition granularity has to be addressed. A fine granularity yields short time periods and provides the means to build a detailed representation of time and time-related activities. But, it results in very large time-space networks and very high solution-method challenges. A coarser granularity alleviates partially this problem, but may result in a poorer representation of decisions and operations in time (Boland et al., 2019). One therefore finds an important body of work, particularly in the context of LTL applications, addressing the issue of finding the appropriate granularity for SSND time-space networks.

The *Dynamic Discretization Discovery* (*DDD*) is a very promising algorithmic strategy addressing the issue through an iterative generation of the time-space network. Introduced by Boland et al. (2017), and enhanced by Hewitt (2019) and Marshall et al. (2021), DDD starts with a coarse granularity and iteratively refines it. The initial *partially time-expanded network* is defined in such a way that the corresponding SSND is a relaxation of the SSND formulated on a time-space network derived from the finest possible granularity in managerial terms (also called “complete enumeration", e.g., the minute). Then, at each iteration of the DDD, the partially time-expanded SSND is solved and the solution is examined to see if it can be converted to an optimal solution to the complete-enumeration SSND, in which case the method stops, Otherwise, the current partially time-expanded network is refined and the algorithm continues. The method performed very well on LTL SSND applications. Successful adaptations to other problem settings were also proposed including, SSND models to determine jointly long-haul and local-delivery routes (Medina et al., 2019; He et al., 2023), and a mixed-integer formulation combining discreet and continuous time representations for a two-echelon location-routing problem, with time-dependent demand and synchronization requirements (Escobar-Vargas and Crainic, 2024).

5 Addressing Resource Management Concerns

As detailed in the previous sections, carriers need physical and human resources to execute services. Resources are generally costly and in short supply, however. Consequently, the resource management and resource-to-service assignment issues are very important for the definition, optimization, and performance of the carrier and its service network.

Traditionally, the literature identified such problems as “operational". More or less sophisticate network flow models are proposed in this context, to be applied over rather short planning horizons, given a tactical plan. None of these approaches works directly on the design of the service network. For detailed surveys and syntheses of the literature, the interested reader may turn to Dejax and Crainic (1987); Cordeau et al. (1998); Piu and Speranza (2014), and the general articles and chapters mentioned in Section 2.

Moreover, many of the early service network design models do not consider resource availability. The goal is a model to provide services with sufficient capacity to transport the demand shipments, assuming that a resource-planning operational model would later be used to provide the resources required to support the selected plan.

Early acknowledgement of the possible shortcomings of such strategies yielded a somewhat more integrative approach, in which the SND formulations account for the need to reposition empty vehicles. Indeed, trade is unbalanced in the nature, volume, and value of the goods exchanged and, therefore, so are the movements of the resources providing the means for transportation. One thus observes surpluses of certain resources at a number of terminals and deficits of the same resources at others. Resource *repositioning*, also called *balancing*, is performed to address this issue and make the system ready for the next round of activities. The method proposed by Crainic et al. (1984) (see also Crainic and Rousseau, 1986; Crainic and Roy, 1988) to account for repositioning within SND formulations computes an origin-destination demand matrix of empty vehicles or power units. This demand, based on the cargo OD-demand matrices and the corresponding estimation of surpluses and deficits at terminals, is then distributed over the network jointly with the cargo demand.

A second integrative approach focuses on the vehicle flows at the local level of each terminal, either by explicitly including decision variables capturing the balancing empty-vehicle numbers to move between terminal pairs (e.g., Powell, 1986a; Smilowitz et al., 2003; Jarrah et al., 2009; Erera et al., 2013), or by integrating the vehicle type into the definition of the service it supports (e.g. Armacost et al., 2002; Lai and Lo, 2004; Pedersen and Crainic, 2007; Bilegan et al., 2022; Kienzle et al., 2024). This strategy is generally identified as *design-balanced*

SND and SSND (e.g. Pedersen et al., 2009). In its original form, it addresses problem settings where each service requires a unit of resource (vehicle) only. The idea is to ensure that the number of inbound services (vehicles) at a node equals the number of outbound services at the same node. The set of node-degree constraints (25) are therefore added to the formulations introduced previously.

$$\sum_{a \in \mathscr{A}_\eta^+} y_a - \sum_{a \in \mathscr{A}_\eta^-} y_a = 0, \ \forall \eta \in \mathscr{N}. \tag{25}$$

This approach may be extended in several ways. First, the service-design variables may be multiplied by an appropriate factor when more than one unit of resource is associated to the service, which ensures the balance of the respective resource at the nodes of the network. Moreover, more than one type of resource may be considered by instantiating a set of constraints similar to (25). Note, however, that additional constraints may be required to govern the inter-resource relations. Consider, to illustrate, several types of container-carrying railcars, with different numbers of platforms and lengths, which can be used simultaneously. Then, one must ensure that the length of the railcar combinations planed to be blocked together or to move together on the same service do not exceed the block or the train maximum permitted length (Kienzle et al., 2024).

Notice that, adding design-balanced constraints to SND formulations complicates the search for high-quality solutions as, for example, even finding an initial solution is no longer straightforward, the rounding of the linear relaxation no longer guaranteeing a feasible solution. Moreover, the size of the formulation is increased by the additional constraints, as is the computational effort to address arc-based models. Heuristic solution methods are therefore generally proposed (see, e.g., Pedersen et al., 2009; Vu et al., 2013; Chouman and Crainic, 2015).

Notice as well that, while models incorporating design-balance type of constraints only acknowledge that service-supporting resources are needed and may have to move empty, they do not recognize that resources need to periodically return to their specific home-base terminal. Moreover, they are not easily extended to account for other considerations such as the size of the fleet, for example. Yet, such constraints induce a structure to the problem and solutions, naturally implying that resources move according to cycles on the (potential) service network.

Cycle-based formulations thus appear natural. These cycles are anchored at the home-base terminal to which the resource is assigned. The cycles may start at different periods during the schedule length and be of different duration, limited in time by the need to return to their home base for inspection and maintenance. Andersen et al. (2009a,b) extend SSND formulations to include resource cycles, cyclic schedules, and the coordination/synchronization of several railroads and navigation services at particular junction points. The authors also show that cycle-based formulations provide more modelling flexibility and computational efficiency. Andersen et al. (2011) exploit this structure in a branch-and-price-based scheme for the problem wherein vehicles flow on cycles and commodities flow on paths, with both cycles and paths generated dynamically via column generation.

Let $\Theta = \{\theta\}$ stand for the set of feasible cycles the units of the resource considered may perform, f_θ the “fixed" cost of selecting and operating the resource cycle $\theta \in \Theta$, and δ_θ^σ the cycle-to-service assignment indicator, where $\delta_\theta^\sigma = 1$ if the resource performing cycle $\theta \in \Theta$ may support service $\sigma \in \Sigma$, and 0 otherwise. Define the binary decision variable $y_\theta = 1$, if cycle

$\theta \in \Theta$ is selected, and 0 otherwise. The basic SSND with single resource cyclic management then becomes

$$\min \sum_{\sigma \in \Sigma} f_\sigma y_\sigma + \sum_{\theta \in \Theta} f_\theta y_\theta + \sum_{k \in \mathscr{K}} \sum_{a \in \mathscr{A}} c_a^k x_a^k \tag{26}$$

subject to constraints (2) - (6) enriched with

$$y_\sigma \leq \sum_{\theta \in \Theta} \delta_\theta^\sigma y_\theta, \ \sigma \in \Sigma, \tag{27}$$

$$y_\theta \in \mathbb{Z}_+, \ \theta \in \Theta, \tag{28}$$

where (26) minimizes the selection and operation costs of services and resources, plus the cost of moving the demand flows, while constraints (27) link the selection of services and the resources required to operate them. Appropriate constraints are added to the formulation to model various *resource management* considerations, such as the cycle duration, unless such characteristics are enforced during the *a priori* or dynamic cycle-generation procedures (Crainic et al., 2014b, 2018a).

Crainic et al. (2014b, 2018a); Hewitt et al. (2019); Crainic and Hewitt (2021) propose more general *Scheduled Service Network Design with Resource Acquisition and Management*, *SSND-RAM*, formulations. The problem settings encompass several types of resources, outsourcing servicing certain markets, as well as resource acquisition, allocation, and re-allocation decisions. The following two-layer (services and resources, Crainic, 2024) SSND-RAM formulation is based on Crainic and Hewitt (2021).

The SSND notation is adjusted for multiple resources. Let $\mathscr{R}$ stand for the set of available resources, f_η^r the fixed cost (salaries, maintenance, etc.) of operating a unit of resource of type $r \in \mathscr{R}$ that is assigned to terminal $\eta \in \mathscr{N}^{\mathrm{PH}}$, and I_η^r the quantity of resources of type r initially assigned to terminal η. Let also Θ_η^r be the set of potential cycles a resource of type r assigned to terminal η can execute, $\Theta^r = \cup_{\eta \in \mathscr{N}^{\mathrm{PH}}} \Theta_\eta^r$, and $\Theta = \cup_{r \in \mathscr{R}} \Theta^r$. The cycle-to-service assignment indicator δ_θ^σ links services and resources as previously. When service costs and capacities vary according to the assigned resource, the notation becomes f_σ^r and $u(\sigma, r)$, $\sigma \in \Sigma$, $r \in \mathscr{R}$, respectively. Notice that a resource-independent fixed service selection cost, f_σ, may still be associated to a service modeling, e.g., the salaries of the officers of a liner ship. Finally, F_σ^r represents the fixed cost of operating service σ with a third party-owned resource of type r.

The service layer of the time-space SSND-RAM network is composed, as previously, of the scheduled (potential) service arcs and holding arcs. The resource layer models the resource cycles, each defined as a sequence of arcs (moving and holding) in the service layer. The resource acquisition and allocation modeling adds a few nodes, $\bar{\mathscr{N}}$, and arcs, $\bar{\mathscr{A}}$, to the resource layer, together with associated parameters and decision variables.

Two types of nodes are included in $\bar{\mathscr{N}}$, symbolically defined at period 0, before the first period of the schedule length. Each node in $\bar{\mathscr{N}}$ is connected to all first representations of the terminal nodes in the resource layer. A unique node is included to represent the acquisition of new resources. We call it A. The arcs $(A, \eta_0) \in \bar{\mathscr{A}}, \eta \in \mathscr{N}^{\mathrm{PH}}$, represent the allocation of newly acquired resources to terminal η at the first period of activity at that terminal (assumed to the first one, for simplicity of presentation). Let h_η^r be the total cost of acquiring a new unit of resource $r \in \mathscr{R}$ and allocating it to terminal $\eta \in \mathscr{N}^{\mathrm{PH}}$.

The second type of node is used to model the re-allocation of existing resources. A node $\bar{\eta}$ is added at period 0 for each terminal $\eta \in \mathcal{N}^{\mathrm{PH}}$, the arcs $(\bar{\eta}, \eta'_0) \in \bar{\mathcal{A}}, \eta' \in \mathcal{N}^{\mathrm{PH}}$, connecting that node to each terminal representing the re-allocation of the resources initially at terminal η to terminal η'. The corresponding cost of repositioning a unit of resource $r \in \mathcal{R}$ from terminal $\eta \in \mathcal{N}^{\mathrm{PH}}$ to terminal $\eta' \in \mathcal{N}^{\mathrm{PH}}$, on arc $(\bar{\eta}, \eta'_0)$, is noted $h^r_{\bar{\eta}\eta'_0}$ (with $h^r_{\bar{\eta}\eta'_0} = 0$ for $\eta = \eta'$).

The cycle definition is extended over the additional nodes and arcs to capture the acquisition and re-allocation activities within the resource-routing decisions. Cycles are thus associated to nodes in $\bar{\mathcal{N}}$ and include the arcs of $\bar{\mathcal{A}}$, yielding the set $\Theta^r_{\bar{\eta}}$ of potential cycles a resource of type r can execute out of each respective terminal $\eta \in \mathcal{N}^{\mathrm{PH}}$.

The decision variables of the SSND, $y_\sigma, \sigma \in \Sigma$, and $x^k_a \geq 0, a \in \mathcal{A}, k \in \mathcal{K}$, are also defined for the SSND-RAM. We define the additional decision variables:

- $y^r_\sigma = 1$, if service $\sigma \in \Sigma$ is operated with a third party-owned resource $r \in \mathcal{R}$, 0, otherwise;
- $z^r_\theta = 1$, if cycle $\theta \in \Theta^r_\eta, r \in \mathcal{R}$, is selected, 0, otherwise;
- $z^r_{\theta\bar{\eta}} = 1$, if cycle $\theta \in \Theta^r_{\bar{\eta}}, \eta \in \mathcal{N}^{\mathrm{PH}}, r \in \mathcal{R}$, is selected, 0, otherwise;
- w^r_η, the number of new units of resource $r \in \mathcal{R}$ acquired and assigned to terminal $\eta \in \mathcal{N}^{\mathrm{PH}}$;
- $w^r_{\bar{\eta}\eta'_0}$, the number of units of resource $r \in \mathcal{R}$ positioned from terminal $\eta \in \mathcal{N}^{\mathrm{PH}}$ to terminal $\eta' \in \mathcal{N}^{\mathrm{PH}}$ (on arc $(\bar{\eta}, \eta'_0)$).

The *Scheduled Service Network Design with Resource Acquisition and Management* formulation for the single-leg-service case may be then written as follows:

$$\text{Minimize} \sum_{r \in \mathcal{R}} \left(\sum_{\eta \in \mathcal{N}^{\mathrm{PH}}} h^r_\eta w^r_\eta + \sum_{(\bar{\eta}, \eta'_0) \in \bar{\mathcal{A}}} h^r_{\bar{\eta}\eta'_0} w^r_{\bar{\eta}\eta'_0} \right) + \tag{29}$$
$$+ \sum_{\sigma \in \Sigma} \left(f_\sigma y_\sigma + \sum_{r \in \mathcal{R}} f^r_\sigma \sum_{\theta \in \Theta^r} \delta^\sigma_\theta z^r_\theta \right) + \sum_{\sigma \in \Sigma} \sum_{r \in \mathcal{R}} F^r_\sigma y^r_\sigma$$
$$+ \sum_{r \in \mathcal{R}} \sum_{\eta \in \mathcal{N}^{\mathrm{PH}}} f^r_\eta \sum_{\theta \in \Theta^r_{\bar{\eta}}} z^r_{\theta\bar{\eta}} + \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}} c^k_a x^k_a$$

$$\text{s.t.} \sum_{(\bar{\eta},\eta'_0)\in\bar{\mathscr{A}}} w^r_{\bar{\eta}\eta'_0} = I^r_\eta,\ r\in\mathscr{R},\ \eta\in\mathscr{N}^{\mathrm{PH}}, \tag{30}$$

$$\sum_{\theta\in\Theta^r_{\bar{\eta}}} z^r_{\theta\bar{\eta}} \leq \sum_{(\bar{\eta},\eta'_0)\in\bar{\mathscr{A}}} w^r_{\bar{\eta}\eta'_0},\ r\in\mathscr{R},\ \bar{\eta}\in\bar{\mathscr{N}}, \tag{31}$$

$$\sum_{a\in\mathscr{A}^+_{\eta_t}} x^k_a - \sum_{a\in\mathscr{A}^-_{\eta_t}} x^k_a = d^k,\ \eta_t\in\mathscr{N}, k\in\mathscr{K}, \tag{32}$$

$$\sum_{k\in\mathscr{K}} x^k_a \leq \sum_{r\in\mathscr{R}} u(\sigma,r)\left(\sum_{\theta\in\Theta^r}\delta^\sigma_\theta z^r_\theta + y^r_\sigma\right),\ a\in\mathscr{A}, \tag{33}$$

$$y_\sigma \leq \sum_{r\in\mathscr{R}}\sum_{\theta\in\Theta^r}\delta^\sigma_\theta z^r_\theta,\ \sigma\in\Sigma, \tag{34}$$

$$y_\sigma + y^r_\sigma \leq 1,\ \sigma\in\Sigma, \tag{35}$$

$$w^r_\eta,\ w^r_{\bar{\eta}\eta'_0}\in\mathbb{Z}^+,\ r\in\mathscr{R}, \bar{\eta}\in\bar{\mathscr{N}}, \eta\in\mathscr{N}^{\mathrm{PH}}, \tag{36}$$

$$z^r_\theta, z^r_{\theta\bar{\eta}}\in\{0,1\},\ r\in\mathscr{R}, \theta\in\Theta^r, \bar{\eta}\in\bar{\mathscr{N}}, \tag{37}$$

$$y^r_\sigma\in\{0,1\},\ \ r\in\mathscr{R},\ \sigma\in\Sigma, \tag{38}$$

$$x^k_a \geq 0,\ a\in\mathscr{A}, k\in\mathscr{K}. \tag{39}$$

The objective minimizes the total cost of the system. The first term models the cost of acquiring new and re-allocating existing resources. The second term computes the cost of selecting services and operating them with owned resources on particular cyclic routes. The third term models the costs incurred to secure third-party resources. The fourth term represents the costs associated with putting a resource into use, while the fifth and last term models shipment transportation costs.

Constraints (30) ensure that all resources of type r that are initially allocated to terminal i are either left at i or re-allocated. Constraints (31) link the strategic resource acquisition and allocation/re-allocation decisions that determine the number of resources available at each terminal with the tactical decision of how many resources from that terminal are to be used to execute services. Note the summation over $\bar{\mathscr{N}}$ in constraints (31) enables the use of resources that are newly acquired.

Constraints (32) and (33) enforce classical network design relations. The former are commodity-specific flow conservation constraints. The latter link the existence of flow on owned or outsourced services to the corresponding service-selection decision. Constraints (34) indicate that at most one resources is used for each owned service, while constraints (35) specify that each service cannot be selected more than once, either supported by the carrier's resources or outsourced. Finally, constraints (36) - (39), define the domains of the variables in the formulation.

The solution methods proposed in Crainic et al. (2014b, 2018a); Hewitt et al. (2019) combine dynamic cycle-generation schemes and mechanisms to choose cycles and services, as well as to move demand flows given the resulting capacity. Moreover, Hewitt et al. (2019) explicitly recognizes uncertainty in demand volume when tactical decisions are taken relative to the design of the service network and the resource allocation (see Section 6).

6 Addressing Uncertainty

The SND and SSND are parameterized mathematical models of consolidation-based transportation systems within the context of planning processes. Using the methodology to support those processes requires not only the model to accurately represent the system, but also the values of the model parameters to adequately predict the variations in the state of the system over the contemplated planning horizon. The models discussed so far are built with deterministic parameter values, which corresponds to assuming the values are certain predictions. Of course, in reality, the validity of this assumption is not certain in most cases. Recall that, tactical plans are actually drawn well in advance of their repetitive execution for a representative, regular operation setting in terms of demand, resource availability, and economic, regulatory, and social environment. Defining a longer schedule length, a week rather that a day, for example, is one way to alleviate the variation of parameter values over the planning horizon, by capturing the corresponding regular, repetitive variations. Thus, for example, instead of a daily value for each OD demand, a week-long schedule length provides the means to model the variations in demand intensity (when appropriate) at each occurrence during the schedule length, with the appropriate estimated value. Yet, the specific details of the shipper requests may change randomly each time they are made and the plan is to be executed. Other parameter values may vary in a similar fashion, including travel times and the cost of outsourcing. Moreover, unexpected events can further influence the execution of scheduled services, altering the planned transportation supply.

Accounting explicitly for uncertainty in SND and SSND models aims to address these issues. Discussing uncertainty and network design in any depth is beyond the scope of this chapter. The interested reader can consult the books by Kall and Wallace (1994); Ruszczyński, A. and Shapiro, A. (2003); Birge and Louveaux (2011); King and Wallace (2012) and Powell (2021) for stochastic programming and uncertainty, and to Hewitt et al. (2021) and Crainic and Rei (2024) for in-depth discussion of uncertainty and network design. We briefly recall the fundamental stochastic programming concepts as applied to service network design in this section.

Uncertainty is generally classified into one of three types based upon their likelihood and impact (Klibi et al., 2010). *Randomness*, refers to events whose likelihood can be described and is reasonably high, but whose impacts can usually be mitigated within normal operations. The classic example of such uncertainty in SND contexts is fluctuations in the shipment volume between a given origin and destination. The second type, *hazards*, refers to events whose likelihood can be described, but are quite rare. An example in SND contexts is infrastructure or vehicle failure. The third type, *deep uncertainty*, refers to events whose likelihood can not be described, but may have a significant impact, e.g., a maritime port closing down due to a threat of terrorist attack. Most of the SND for consolidation-based carriers research has focused on the first type of uncertainty, randomness, particularly with respect to model parameter values. This uncertainty is modeled by extending one of these deterministic models to a stochastic-programming formulation. We reflect this state-of-the-art.

The challenge when planning within an uncertain context raises from the combination of the timing of decisions and the information available at decision time. Thus, tactical plans, which include the services, resources, and commodity itineraries selected, are built at a time when information regarding the values of the SND parameters is incomplete and only statistical distributions are known. Then, each time the plan is to be executed, the actual values of those parameters are *revealed*, or *observed*, and become known. Even the most well-conceived plan may then become infeasible or financially burdensome to implement because of the often-costly significant revisions it requires. The critical question, then, is how to effectively address this challenge within a SND model supporting the initial tactical planning phase performed by carriers.

There are two general stochastic-optimization modeling paradigms applied to SND. The first involves imposing a set of probabilistic constraints, known as *chance constraints*, to limit the likelihood of solutions becoming infeasible at execution time. *Stochastic models with recourse* follow the second paradigm, which explicitly identifies the decisions in accordance with the *information-revelation* process inherent to the problem. The models thus explicitly differentiate between decisions made under uncertainty, when the plan is built, referred to as *a priori* decisions, and those made after the uncertainty is revealed, when the plan is to be revised and then executed, referred to as *recourse* decisions.

We present these two paradigms as they are applied to demand uncertainty and the basic linear-cost SND model (see Hewitt et al., 2021; Crainic and Rei, 2024, for discussion of uncertainty related to other paramers). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be the probability space associated with a random experiment that reveals the demand volumes $d^k, k \in \mathscr{K}$. The set Ω contains the possible outcomes $\omega \in \Omega$ of the random experiment, $\mathscr{F}$ defines the set of events, and $\mathbb{P}$ is the measure assigning probabilities to the possible outcomes of the random experiment.

6.1 Chance Constraints

Two cases may be contemplated when demand uncertainty is considered. On the one hand, the designed service network and commodity itineraries are inadequate to cover the total demand issued by the shippers. Such circumstances often compel carriers to *outsource* moving some demand, that is, to resort to *ad-hoc capacity*, incurring substantially higher costs or penalties than operating their own resources and services. On the other hand, there's the risk that selected services may not support the intended commodity flow to be transported due to unforeseen reductions in available capacities, undermining the feasibility of the planned itineraries. Chance constraints provide the means to embed these probabilistic considerations within the tactical planning framework, offering a method to preemptively address potential execution challenges and thus mitigate risk.

Recall the dummy arcs $a^k = (O(k), D(k))$ for each commodity $k \in \mathscr{K}$ introduced in Section 2.4 to model the utilization of ad-hoc capacity for transporting overflowing portions of commodity volumes. Given the uncertainty on demand volumes and the resulting uncertainty on the need for outsourcing, let $\varsigma^k(\omega)$, $\omega \in \Omega, k \in \mathscr{K}$, be the *realization-specific slack-flow variables*. The probabilistic constraints imposed on the tactical plans can then be formulated by setting limits on the flow values assigned to these variables.

Following Hewitt et al. (2021) and Crainic and Rei (2024), let $F^k_\eta(x) \sum_{a \in \mathscr{A}^+_\eta} x^k_a - \sum_{a \in \mathscr{A}^-_\eta} x^k_a$ be the net-flow function for commodity $k \in \mathscr{K}$ at node $\eta \in \mathscr{N}$. A chance constraint that limits the total amount observed on the slack-flow variables can then be defined as follows:

$$\mathbb{P}\left(\omega \in \Omega \mid \exists \varsigma^k(\omega) \geq 0 : \begin{array}{c} \sum\limits_{k \in \mathscr{K}} \varsigma^k(\omega) \geq \alpha_{\mathscr{K}}, \\ F^k_{O(k)}(x) = d^k(\omega) - \varsigma^k(\omega), \\ F^k_{D(k)}(x) = -d^k(\omega) + \varsigma^k(\omega), \end{array} \right) \leq 1 - \beta_{\mathscr{K}},\ k \in \mathscr{K}. \qquad (40)$$

The chance constraint (40) limits the probability of observing realizations $\omega \in \Omega$ where the total commodity volumes that are transported on ad-hoc capacity exceeds the threshold $\alpha_{\mathscr{K}}$ (i.e., $\sum_{k \in \mathscr{K}} \varsigma^k(\omega) \geq \alpha_{\mathscr{K}}$) to at most $1 - \beta_{\mathscr{K}}$, where $\beta_{\mathscr{K}}$ stands for the requisite level of reliability for the planned itineraries. Such conditions may also be established on a per-commodity basis, by defining a commodity-specific threshold $\alpha_k, k \in \mathscr{K}$ to indicate the limit beyond which the flow assigned to $\varsigma^k(\omega)$ is considered as an unacceptable outcome $\omega \in \Omega$.

6.2 Stochastic Models with Recourse

Stochastic SND models with recourse make up the alternate stochastic-programming paradigm for addressing uncertainty in consolidation-based freight transportation planning. These models are designed to reflect the decision-making dynamics of the carrier tactical planning and execution processes. Specifically, a stochastic formulation comprehensively incorporates both tactical decisions for designing the service network and operational-level decisions for the repeated adjustment and application of the tactical plan, once the values of the random parameters are revealed. Then, the objective of the stochastic SND model is to devise a plan that "optimizes" not only the unchanging part of the plan, but also the expected cost over the planning horizon associated to repeatedly adjusting and executing it.

This setting inherently encompasses multiple *stages*, that is, specific moments in time when stochastic parameters become known, allowing decisions to be made in reaction to this newly acquired information. In a general sense, the tactical decisions are thus made during the *first stage*, often referred to as *a priori* decisions, under complete uncertainty. Meanwhile, operational decisions are naturally executed in subsequent stages, known as *recourse* decisions, as the stochastic parameters become progressively observed.

Stochastic multi-stage formulations provide the means for a more refined representation of the dynamics of operational decision-making. They also, however, considerably increase the complexity of the optimization model, especially as it expands with additional stages and when discreet decisions are part of some of the later stages. Moreover, detailed operational specifics may not be essential for accurately establishing tactical plans, a good approximation being often sufficient to provide guidance. It is often the case, for example, that those later-stage decisions are included in the tactical plan (e.g., demand itineraries) but are not expected to be implemented. They are rather to be determined for the revealed values each time the plan is executed or, at best, to guide the adjustment of the plan. Their role then is to approximate the impact of the first stage decisions on the performance of the system over the planning horizon. Consequently, the predominant modeling approach in the SND literature employs a two-stage formulation.

How each stage is defined should reflect the specific planning problem under consideration. Typically, the first stage involves design decisions that specify the services to be operated, as well as the resource-management decisions when relevant, thereby establishing the carrier service network. These decisions are made well before the actual utilization of the services, requiring them to comprehensively address the full spectrum of anticipated uncertainty.

A broad range of potential recourse actions can be established for the second stage, representing varying degrees of flexibility in adjusting the tactical plan. On one end of the spectrum is the *simple recourse* option, which involves imposing a penalty proportional to the extent of the plan's infeasibility. This approach essentially adopts a general *observe-and-pay* strategy for defining recourse actions. More complex recourse actions are formulated at the network level. A notable strategy within this context is the establishment of demand itineraries as second-stage decisions, effectively treating the flow decisions as recourse actions. This is the *network strategy* taken in the illustrative model presented below and has been shown to be effective in a variety of tactical planning transportation problems (e.g., Lium et al., 2007, 2009; Hoff et al., 2010; Crainic et al., 2011, 2014a; Bai et al., 2014; Dong et al., 2015; Demir et al., 2016; Crainic et al., 2016; Hasany and Shafahi, 2017; Lanza et al., 2018; Wang et al., 2019; Hewitt et al., 2019; Lanza et al., 2021; Scherr et al., 2022; Jiang et al., 2021; Liu et al., 2023). Recourse actions can also extend to structural adjustments in the service network initially designed during the first stage (e.g., Crainic et al., 2016; Müller et al., 2021). However, adopting such a strategy introduces greater complexity, as it requires making discrete decisions as part of the recourse actions.

In all cases, as the *second stage / recourse* decisions are functions of random variables, they are random variables as well. Moreover, the general two-stage modeling approach assumes that all uncertainty is clarified in a single stage, meaning all stochastic parameters are assumed to be observed simultaneously. Therefore, the second stage offers an approximation of the carrier's real operations to meet shipper demands, guided by the tactical plan established in advance. Consequently, the objective of the model is to minimize the sum of the costs associated with the first stage decisions and the expected costs associated with second stage decisions. Solving the two-stage stochastic model thus yields a comprehensive tactical plan, which includes the setup of the service network, as well as the cost-efficient adjustments to accommodate observed random changes.

Surveying the literature on stochastic service network design, one notices that most models prescribe the selection of services in the first stage and the routing of commodities, given those services and the realized parameter values, in the second stage. We therefore present such a formulation to illustrate the previous discussion. Let the variables y be the first-stage design decisions, and $x_a^k(\omega), a \in \mathscr{A}, k \in \mathscr{K}, \omega \in \Omega$, be the flow variables representing the network recourse. Assuming a linear definition of the system cost, as in equation (1), the model formulation is (41) - (47):

$$\min \quad \sum_{\sigma \in \Sigma} f_\sigma y_\sigma + \mathbb{E}_\xi \Big[Q\big(y, \xi(\omega)\big) \Big] \tag{41}$$

$$\text{s.t.} \qquad y_\sigma \in \mathbb{Z}_+, \sigma \in \Sigma, \tag{42}$$

where,

$$Q\big(y,\xi(\omega)\big) = \min \sum_{k\in\mathscr{K}} \sum_{a\in\mathscr{A}} c_a^k x_a^k(\omega) \tag{43}$$

$$\text{s.t.} \sum_{a\in\mathscr{A}_\eta^+} x_a^k(\omega) - \sum_{a\in\mathscr{A}_\eta^-} x_a^k(\omega) = d_k(\omega), \qquad \eta \in \mathscr{N}, k \in \mathscr{K}, \tag{44}$$

$$\sum_{k\in\mathscr{K}} x_a^k(\omega) \le u_a y_{\sigma(a)}, \qquad a \in \mathscr{A}, \tag{45}$$

$$x_a^k(\omega) \le u_a^k y_{\sigma(a)}, \qquad a \in \mathscr{A}, k \in \mathscr{K}, \tag{46}$$

$$x_a^k(\omega) \ge 0, \qquad a \in \mathscr{A}, k \in \mathscr{K}, \tag{47}$$

where $d_k(omega) = \text{vol}_k(\omega)$ at the demand origin $\eta = O(k)$, equals $-\text{vol}_k(\omega)$ at the destination node of the demand $\eta = D(k)$, and zero at all other nodes.

Model (41) - (42) addresses the first-stage decisions design for the service network under complete uncertainty. The objective function (41) combines two components. The first, deterministic component computes the total fixed costs for the selected services. The second is the recourse cost function representing an aggregated (the expectation) measure of the anticipated future costs of adjusting / adapting the tactical plan, given by the design decisions, to revealed information. The recourse cost function $Q\big(y,\xi(\omega)\big)$, defined for the realization $\omega \in \Omega$, evaluates the total cost incurred to optimally move the realized demand $\xi(\omega) = \text{vol}_k(\omega)$ given the service network y defined in the first stage. The function $Q\big(y,\xi(\omega)\big)$ is thus characterized as the optimal solution to the minimum cost multi-commodity flow model (43)-(47).

7 Conclusions and Perspectives

The chapter presented an overview and synthesis of the main classes of Service Network Design models aimed at supporting decision-making in planning the activities and managing the resources of consolidation-based freight carriers and systems. The chapter focused on issues and model structures of general interest and relevance. We continue this approach in identifying a number of challenging research perspectives of importance for service network design and its applications.

Extending the scope of the SND methodology, with its related modeling challenges, makes up a first broad and important research field. The aim is to enhance the representation capability and relevance of our models, and extend the applicability of SND methodology to other planning levels and applications. As always in Operations Research, one must start with a proper definition and specification of the problem under study, then proceed to model the system and decisions adequately, before developing the appropriate solution method. The three phases require a significant research effort. It is also noticeable that advances in general methodology are often prompted by model and method developments addressing challenges in particular problem settings. The research trajectory from particular applications, to general methodology, to tailoring the enhanced methodology to different applications settings is a challenging but highly beneficial characteristic of our profession.

We have discussed the integration of resource-management concerns in SND models, but many challenges remain. First, enhance the modeling of major work rules, such as the maintenance periods for equipment, and particular qualifications and permissions to operate various equipment types for crews. Second, irrespective of mode and setting, transportation services require in many cases resources of several types, each with its own particular operational, maintenance, and fleet-size characteristics, the combinations being governed by particular compatibility rules to operate. Third, the integration of management concerns for several heterogeneous interlinked resources into SND and SSND formulations challenges the modeling of the relations among resources, and among them and the services they may support. The resulting models are also significantly larger, particularly when the time dimension is also explicitly considered, requiring the development of appropriate efficient solution methods.

Explicitly addressing time and delay-related issues enlarges and refines the scope of SND models while raising significant modeling and algorithmic challenges. A first major research area is the proper definition of time as integrated into SSND formulations. The granularity issue is still not comprehensively addressed. On the one side, there is the issue of adequately identifying the partition of the schedule length at various terminals and whether there is the need for a uniform partition. Then, one must face the issue of an efficient time-space network definition given strict service schedules or when particular (individual) temporal representations of the system's terminals are contemplated. The definition of mini-terminal time-space models and their linkage to service schedules is both a beneficial and a challenging modeling development. A second research avenue in this area is the generalization of such studies to the multi-resource problem settings, which give raise to multi-layer networks. The development of efficient solution methods is required in all cases. The enhancement and generalization of the Dynamic Discretization Discovery approach makes up a promising direction, as is its combination with meta- and matheuristics able to address large problem instances.

Changing the modeling paradigm and considering time implicitly in SND models is a long-time goal in the research community. A few recent contributions (e.g., He et al., 2023; Lanza et al., 2024) open interesting perspectives that invite for a sustained research effort in the area.

Considering time, be it explicitly or implicitly, conducts one to study the duration of activities, particularly in terminals and particularly when several activities compete for the same restricted capacity. We discussed some of these delay-related issues and the models proposed in the literature. Yet, this field is still very little explored and many interesting research issues are open, including the more refined modeling of mini-terminals and the activities herein, and the treatment of delays in terms of model representativity and algorithmic efficiency. Related to the latter is the issue of approximating the delays with linear or non-linear, ideally convex, functions. The former makes for an easier algorithm development, while the latter offers a more refined and adequate representation. More generally, research is needed into network design formulations with non-linear objective functions and the associated solution methods.

Addressing uncertainty in transportation planning and SND models and methods constitutes a broad and important research area, challenging modeling and algorithmic development alike. We discussed two modeling approaches and mentioned several contributions to the field. One may state, however, that research in uncertainty and SND is still in its infancy. Research is still required in adequately representing demand uncertainty in the various problem settings.

Almost totally overlooked, although of great operational and economic importance, is the uncertainty in travel and terminal-activity times. The solution often adopted in practice of adding large buffers to the planned delivery times is not only scientifically unsatisfactory, but also less and less economically viable and impracticable in many cases (City Logistics and Synchromodality to name but two examples). Moreover, one should not overlook that both demand and time uncertainty (and heavy correlations) characterize operations and their simultaneous presence and interactions, should be reflected in the planning models proposed. Research on this challenging topic is needed.

As indicated, the discussion and model approaches of Section 6 refer to business-as-usual cases, when uncertainty can be somewhat easily represented with probability distributions. Other sources of uncertainty exist, however, and should be studied. Reliability and robustness are two such issues, as is resilience, i.e., the capability to rebound following an incident, and the operation plans to perform the recovery and return to a desired state of system and operation behavior. Advancing in this direction would also lead to a broader exploration of information-revelation mechanisms and multi-stage formulations.

We complete this “modeling" discussion by returning to the overture of the section and the need to extend the field of interest and development. Multi-stakeholder problem settings are a first research area in such respect. These problems typically include multiple stakeholders (e.g., carriers) that mutualize and coordinate their resources, either directly or through some form of intermediaries (Taherkhani et al., 2022; Bruni et al., 2024), to support and execute the services that constitute the overall transportation supply. The accurate definition of the stakeholder interactions, with the associated cooperation mechanisms, and their modeling within SND formulations make up a major research challenge.

Multi-Layered SSND problems, which involve planning services across interconnected and layered systems, are also gaining attention. Yet, planning services over multi-layered systems presents significant discrete optimization challenges. Such systems often encompass diverse transportation operations being conducted over varying distances, and incorporate different transportation modes both traditional and environmentally friendly options. Each layer and mode thus introduces unique complexities that must be addressed effectively. Furthermore, the design decisions in some layers are defined in terms of sets of design decisions in other layers, being connected through complex relationships (Crainic, 2024). This complexity results in significantly challenging discrete optimization models that are large-scale, possibly non-linear, and involve elements of uncertainty to be solved.

It is worth noticing that most SND models discussed in this chapter assume that the behavior of customers, that is, of demand, is known with respect to economic, e.g., tariffs, and service-level criteria. This is true even when uncertainty in these elements is explicitly represented. Or, customers do react to tariffs and require quality-of-service levels and, consequently, so is the demand the carrier will ultimately service and the revenues it can potentially earn. Extending the SND to address such issues requires considering not only a profit-maximizing objective, but also modeling in mathematical terms the behavioral relations between tariffs, service-quality levels, and the willingness of customers to give a carrier their business. The revenue management literature is the starting point of this line of research noticing, however, that most of it targets people-servicing industries and that one cannot simply transpose those results to the freight transport environment (Bilegan et al., 2022).

Addressing large SND models makes up an extremely important but challenging research area. Space is lacking to explore this area in any comprehensive way. Hence, we only mention a few particularly interesting avenues (we mentioned already the need to explore solution methods when time is explicitly modeled): 1) Dynamic generation of services (paths), resource work assignments (cycles), blocks (paths), and demand-flow itineraries (paths), including extending the column-generation methodology to the SND and SSND cases with simultaneous generation of several types of paths and cycles; 2) Recalling that network design problems are NP-Hard in most cases of interest, and service network design ones are not different, research is needed in heuristic-type solution methods, matheuristics, in particular, which combine exact and meta-heuristic solution principles, ideally coupled with parallel optimization strategies, such as the Integrative Cooperative Search (Crainic, 2019); 3) Development of efficient solution methods for stochastic SND and SSND, which are particularly challenging, even for the two-stage formulations of business-as-usual demand uncertainty case, which has been studied the most. Efficient decomposition and scenario-reduction methods (e.g., Rahmaniani et al., 2017; Hewitt et al., 2021, 2022) offer interesting starting points for what should be a significant research effort on exact and matheuristic solution methods for stochastic service network design.

Acknowledgements

While working on the project, the author held the UQAM Chair on Intelligent Logistics and Transportation Systems Planning, and was Adjunct Professor, Department of Computer Science and Operations Research, Université de Montréal. We gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery grant program. We also gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants.

References

Ahuja, R.K., Cunha, C.B., and Şahin, G. (2005a). Network Models in Railroad Planning and Scheduling. In *Tutorials in Operations Research INFORMS 2005*, pages 54–101. INFORMS.

Ahuja, R.K., Liu, J., Orlin, J.B., and Sharma, D.and Shughart, L.A. (2005b). Solving Real-Life Locomotive-Scheduling Problems. *Transportation Science*, 39(4):503–517.

Albinski, S., Crainic, T.G., and Minner, S. (2020). The Day-before Truck Platooning Planning Problem and the Value of Autonomous Driving. Technical Report CIRRELT-2020-04, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada.

Ambra, T.K., Caris, A., and Macharis, C. (2019). Towards Freight Transport System Unification: Reviewing and Combining the Advancements in the Physical Internet and Synchromodal Transport Research. *International Journal of Production Research*, 57(6):1606–1623.

Ambra, T.K., Mommens, C., and Macharis, C. (2021). Intermodal and Synchromodal Freight Transport. In Vickerman, R. et al., editor, *International Encyclopedia of Transportation*, volume 3, pages 456–462. Elsevier U.K.

Ammann, P., Albinski, S., Crainic, T.G., and Kolisch, R. (2024). Joint Truck and Driver Routing and Scheduling for Semi-Autonomous Platoons. Technical Report CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada.

Andersen, J. and Christiansen, M. (2009a). Designing New Europeaan Rail Freight Services. *Journal of the Operational Research Society*, 60:348–360.

Andersen, J. and Christiansen, M. (2009b). Designing new european rail freight services. *Journal of the Operational Research Society*, 60(January):348–360.

Andersen, J., Christiansen, M., Crainic, T.G., and Grønhaug, R. (2011). Branch-and-Price for Service Network Design with Asset Management Constraints. *Transportation Science*, 46(1):33–49.

Andersen, J., Crainic, T.G., and Christiansen, M. (2009a). Service Network Design with Asset Management: Formulations and Comparative Analyzes. *Transportation Research Part C: Emerging Technologies*, 17(2):197–207.

Andersen, J., Crainic, T.G., and Christiansen, M. (2009b). Service Network Design with Management and Coordination of Multiple Fleets. *European Journal of Operational Research*, 193(2):377–389.

Armacost, A.P., Barnhart, C., and Ware, K.A. (2002). Composite Variable Formulations for Express Shipment Service Network Design. *Transportation Science*, 36(1):1–20.

Assad, A.A. (1980). Models for Rail Transportation. *Transportation Research Part A: Policy and Practice*, 14:205–220.

Bai, R., Wallace, S. W., Li, J., and Chong, A. Y.-L. (2014). Stochastic Service Network Design with Rerouting. *Transportation Research Part B: Methodological*, 60:50–65.

Bakir, I., Erera, A., and Savelsbergh, M.W.F. (2021). Motor Carrier Service Network Design. In Crainic, T.G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications in Transportation and Logistics*, chapter 16, pages 427–467. Springer, Boston.

Balakrishnan, A., Kuo, A., and Si, X. (2016). Real-Time Decision Support for Crew Assignment in Double-Ended Districts for U.S. Freight Railways. *Transportation Science*, 50(4):1139–1393.

Ballot, E., Montreuil, B., and Meller, R. D. (2014). *The Physical Internet: The Network of Logistics Networks*. La Documentation Française, Paris, France.

Bauer, J., Bektaş, T., and Crainic, T.G. (2010). Minimizing Greenhouse Gas Emissions in Intermodal Freight Transport: An Application to Rail Service Design. *Journal of the Operational Research Society*, 61:530–542.

Bektaş, T. and Crainic, T.G. (2008). A Brief Overview of Intermodal Transportation. In Taylor, G.D., editor, *Logistics Engineering Handbook*, chapter 28, pages 1–16. Taylor and Francis Group, Boca Raton, FL, USA.

Bektaş, T., Crainic, T.G., and Van Woensel, T. (2017). From managing urban freight to smart city logistics networks. In Gakis, K. and Pardalos, P., editors, *Networks Design and Optimization for Smart Cities*, volume 8 of *Series on Computers and Operations Research*, pages 143–188. World Scientific Publishing.

Bilegan, I.C., Brotcorne, L., Feillet, D., and Hayel, Y. (2015). Revenue Management for Rail Container Transportation. *EURO Journal on Transportation and Logistics*, 4(2):261–283.

Bilegan, I.C., Crainic, T.G., and Wang, Y. (2022). Scheduled Service Network Design with Revenue Management Considerations and an Intermodal Barge Transportation Illustration. *European Journal of Operational Research*, 300(1):164–177.

Birge, J.R. and Louveaux, F. (2011). *Introduction to Stochastic Programming, 2nd Edition*. Springer, New York.

Boland, N., Hewitt, M., Marshall, L., and Savelsbergh, M.W.F. (2017). The Continuous-Time Service Network Design Problem. *Operations Research*, 65(5):1303–1321.

Boland, N., Hewitt, M., Marshall, L., and Savelsbergh, M.W.F. (2019). The Price of Discretizing Time: A Study in Service Network Design. *EURO Journal on Transportation and Logistics*, 8(2):195–216.

Bruni, M.E., Crainic, T.G., and Perboli, G. (2024). Multilayer network design for consolidation-based transportation planning. In Crainic, T.G., Gendreau, M., and A., F., editors, *Contributions to Combinatorial Optimization and Applications - A Tribute to Bernard Gendron*. Springer. forthcoming.

Chouman, M. and Crainic, T.G. (2015). Cutting-Plane Matheuristic for Service Network Design with Design-Balanced Requirements. *Transportation Science*, 49(1):99–113.

Chouman, M. and Crainic, T.G. (2021). Freight Railroad Service Network Design. In Crainic, T.G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications in Transportation and Logistics*, chapter 13, pages 383–426. Springer, Boston.

Christiansen, M., Fagerholt, K., Nygreen, B., and Ronen, D. (2007). Maritime Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, pages 189–284. North-Holland, Amsterdam.

Christiansen, M., Fagerholt, K., and Ronen, D. (2004). Ship Routing and Scheduling: Status and Perspectives. *Transportation Science*, 38(1):1–18.

Christiansen, M., Helsen, E., Pisinger, D., Sacramento, D., and Vilhelmsen, C. (2020). Liner Shipping Network Design. *European Journal of Operational Research*, 286(1):1–20.

Christiansen, M., Helsen, E., Pisinger, D., Sacramento, D., and Vilhelmsen, C. (2021). Liner Shipping Network Design. In Crainic, T.G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications in Transportation and Logistics*, chapter 15, pages 469–505. Springer, Boston.

Cordeau, J.F., Toth, P., and Vigo, D. (1998). A Survey of Optimization Models for Train Routing and Scheduling. *Transportation Science*, 32(4):380–404.

Crainic, T.G. (1988). Rail Tactical Planning: Issues, Models and Tools. In Bianco, L. and La Bella, A., editors, *Freight Transport Planning and Logistics*, pages 463–509. Springer-Verlag, Berlin.

Crainic, T.G. (2000). Network Design in Freight Transportation. *European Journal of Operational Research*, 122(2):272–288.

Crainic, T.G. (2003). Long-Haul Freight Transportation. In Hall, R.W., editor, *Handbook of Transportation Science*, pages 451–516. Kluwer Academic Publishers, Norwell, MA, second edition.

Crainic, T.G. (2009). Service Design Models for Rail Intermodal Transportation. In Bertazzi, L., Speranza, M.G., and van Nunen, J.A.E.E., editors, *Lecture Notes in Economics and Mathematical Systems*, volume 619, pages 53–67. Springer-Verlag, Berlin Heidelberg.

Crainic, T.G. (2019). Parallel Metaheuristics and Cooperative Search. In Gendreau, M. and Potvin, J.-Y., editors, *Handbook of Metaheuristics (3d Edition)*, pages 419–451. Springer.

Crainic, T.G. (2024). Multi-layer network design for consolidation-based transportation planning. In Crainic, T.G., Gendreau, M., and A., F., editors, *Contributions to Combinatorial Optimization and Applications - A Tribute to Bernard Gendron*, chapter 9. Springer. forthcoming.

Crainic, T.G., Errico, F., Rei, W., and Ricciardi, N. (2016). Modeling Demand Uncertainty in Two-Tier City Logistics Tactical Planning. *Transportation Science*, 50(2):559–578.

Crainic, T.G., Ferland, J.-A., and Rousseau, J.-M. (1984). A Tactical Planning Model for Rail Freight Transportation. *Transportation Science*, 18(2):165–184.

Crainic, T.G. and Florian, M. (2008). National Planning Models and Instruments. *INFOR*, 46(4):81–90.

Crainic, T.G., Florian, M., Guélat, J., and Spiess, H. (1990a). Strategic Planning of Freight Transportation: STAN, An Interactive-Graphic System. *Transportation Research Record*, 1283:97–124.

Crainic, T.G., Florian, M., and Léal, J.-E. (1990b). A Model for the Strategic Planning of National Freight Transportation by Rail. *Transportation Science*, 24(1):1–24.

Crainic, T.G., Fu, X., Gendreau, M., Rei, W., and Wallace, S.W. (2011). Progressive Hedging-based Meta-heuristics for Stochastic Network Design. *Networks*, 58(2):114–124.

Crainic, T.G. and Gendreau, M. (1986). Approximate Formulas for the Computation of Connection Delays under Capacity Restrictions in Rail Freight Transportation. In *Research for Tomorrow's Transport Requirements*, volume 2, pages 1142–1155. Fourth World Conference on Transport Research, Vancouver, Canada.

Crainic, T.G., Gendreau, M., and Gendron, B. (2021a). Fixed-Charge Network Design Problems. In Crainic, T.G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications in Transportation and Logistics*, chapter 2, pages 15–28. Springer, Boston.

Crainic, T.G., Gendreau, M., and Gendron, B., editors (2021b). *Network Design with Applications in Transportation and Logistics*. Springer, Boston.

Crainic, T.G., Gendron, B., and Kazemzadeh, M. (2022). A Taxonomy of Multilayer Network Design and a Survey of Transportation and Telecommunication Applications. *European Journal of Operational Research*, 303(1):161–179.

Crainic, T.G.., Gonzalez-Feliu, J., Ricciardi, N., Semet, F., and Van Woensel, T. (2023a). Operations research for planning and managing city logistics systems. In Marcucci, E., Gatta, V., and Le Pira, M., editors, *Handbook on City Logistics and Urban Freight*, chapter 10, pages 190–223. Edward Elgar, Northampton, MA, U.S.A,.

Crainic, T.G.., Gonzalez-Feliu, J., Ricciardi, N., Semet, F., and Van Woensel, T. (2023b). Operations research for planning and managing city logistics systems. In Marcucci, E., Gatta, V., and Le Pira, M., editors, *Handbook on City Logistics and Urban Freight*, chapter 10, pages 190–223. Edward Elgar, Northampton, MA, U.S.A.,.

Crainic, T.G. and Hewitt, M. (2021). Service Network Design. In Crainic, T.G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications in Transportation and Logistics*, chapter 12, pages 347–382. Springer, Boston.

Crainic, T.G., Hewitt, M., and Rei, W. (2014a). Scenario Grouping in a Progressive Hedging-based Meta-heuristics for Stochastic Network Design. *Computers & Operations Research*, 43(2):90–99.

Crainic, T.G., Hewitt, M., Toulouse, M., and Vu, D.M. (2014b). Service Network Design with Resource Constraints. *Transportation Science*, 50(4):1380–1393.

Crainic, T.G., Hewitt, M., Toulouse, M., and Vu, D.M. (2018a). Scheduled Service Network Design with Resource Acquisition and Management. *EURO Journal on Transportation and Logistics*, 7(3):277–309.

Crainic, T.G. and Kim, K.H. (2007). Intermodal Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter 8, pages 467–537. North-Holland, Amsterdam.

Crainic, T.G. and Laporte, G. (1997). Planning Models for Freight Transportation. *European Journal of Operational Research*, 97(3):409–438.

Crainic, T.G. and Montreuil, B. (2016). Physical Internet Enabled Hyperconnected City Logistics. In Taniguchi, E. and Thompson, R. G., editors, *9th International Conference on City Logistics, June 2015, Tenerife, Spain*, volume 12 of *Transportation Research Procedia*, pages 383–398. Elsevier.

Crainic, T.G., Perboli, G., and Ricciardi, N. (2021c). City Logistics. In Crainic, T.G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications in Transportation and Logistics*, chapter 16, pages 507–537. Springer, Boston.

Crainic, T.G., Perboli, G., and Rosano, M. (2018b). Simulation of Intermodal Freight Transportation Systems: A Taxonomy. *European Journal of Operational Research*, 270:401–418.

Crainic, T.G. and Quérin, L. (1988). Approximation of Over-the-line Delay for Single Line - Single Speed Freight Rail Transportation. Publication 632, Centre de recherche sur les transports, (CRT) Université de Montréal, Montreal, Canada.

Crainic, T.G. and Rei, W. (2024). 50 Years of Operations Research for Planning Consolidation-based Freight Transportation. Publication CIRRELT-2024, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Université de Montréal, Montreal.

Crainic, T.G., Ricciardi, N., and Storchi, G. (2009). Models for Evaluating and Planning City Logistics Transportation Systems. *Transportation Science*, 43(4):432–454.

Crainic, T.G. and Rousseau, J.-M. (1986). Multicommodity, Multimode Freight Transportation: A General Modeling and Algorithmic Framework for the Service Network Design Problem. *Transportation Research Part B: Methodological*, 20:225–242.

Crainic, T.G. and Roy, J. (1988). O.R. Tools for Tactical Freight Transportation Planning. *European Journal of Operational Research*, 33(3):290–297.

Crainic, T.G. and Roy, J. (1992). Design of Regular Intercity Driver Routes for the LTL Motor Carrier Industry. *Transportation Science*, 26(4):280–295.

Crainic, T.G.., W., K., and Montreuil, B. (2023c). Hyperconnected city logistics: A conceptual framework. In Marcucci, E., Gatta, V., and Le Pira, M., editors, *Handbook on City Logistics and Urban Freight*, chapter 20, pages 398–421. Edward Elgar, Northampton, MA, U.S.A.,.

Croxton, K.L., Gendron, B., and Magnanti, T.L. (2003a). A Comparison of Mixed-integer Programming Models for Nonconvex Piecewise Linear Cost Minimization Problems. *Management Science*, 49(9):1268—-1273.

Croxton, K.L., Gendron, B., and Magnanti, T.L. (2003b). Models and Methods for Merge-in-Transit Operations. *Transportation Science*, 37(1):1–22.

Croxton, K.L., Gendron, B., and Magnanti, T.L. (2007). Variable Disaggregation in Network Flow Problems with Picewise Linear Costs. *Operations Research*, 55(1):146–157.

Dejax, P.J. and Crainic, T.G. (1987). A Review of Empty Flows and Fleet Management Models in Freight Transportation. *Transportation Science*, 21(4):227–247.

Delorme, L., Roy, J., and Rousseau, J.-M. (1988). Motor-Carrier Operation Planning Models: A State of the Art. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics*, pages 510–545. Springer-Verlag, Berlin.

Demir, E., Burgholzer, W., Hrušovský, M., Arikan, E., Jammernegg, W., and Woensel, T. V. (2016). A Green Intermodal Service Network Design Problem with Travel Time Uncertainty. *Transportation Research Part B: Methodological*, 93:789–807.

Dong, J.-X., Lee, C.-Y., and Song, D.-P. (2015). Joint Service apacity Planning and Dynamic Container Routing in Shipping Network with Uncertain Demands. *Transportation Research Part B: Methodological*, 78:404–421.

Erera, A., Hewitt, M., M,.W.F., S., and Zhang, Y. (2013). Improved Load Plan Design Through Integer Programming Based Local Search. *Transportation Science*, 47(3):412–427.

Escobar-Vargas, D. and Crainic, T.G. (2024). Multi-attribute Two-Echelon Location Routing: Formulations and Dynbamic Discretization Discovery Approach. *European Journal of Operational Research*, 314(1):66–78.

Fontaine, P., Crainic, T.G., Jabali, O., and Rei, W. (2021). Scheduled Service Network Design with Resource Management for Two-tier Multimodal City Logistics. *European Journal of Operational Research*, 294(2):558–570.

Ford, L.R. and Fulkerson, D.R. (1958). Constructing Maximal Dynamic Flows from Static Flows. *Operations Research*, 6(3):419–433.

Giusti, R., Manerba, D., and Tadei, R. (2019). Synchromodal Logistics: An Overview of Critical Success Factors, Enabling Technologies, and Open Research Issues. *Transportation Research Part E: Logistics and Transportation Review*, 129:92–110.

Guélat, J., Florian, M., and Crainic, T.G. (1990). A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows. *Transportation Science*, 24(1):25–39.

Hasany, R.M. and Shafahi, Y. (2017). Two-stage Stochastic Programming for the Railroad Blocking Problem with Uncertain Demand and Supply Resources. *Computers & Industrial Engineering*, 106:275–286.

He, Y., Hewitt, M., Lehuédé, F., Medina, N., and Péton, O. (2023). A Continuous-time Service Network Design and Vehicle Routing Problem. *Networks*, 83(2):300–323.

Hewitt, M. (2019). Enhanced Dynamic Discretization Discovery for the Continuous-Time Load Plan Design Problem. *Transportation Science*, 53(6):1731–1750.

Hewitt, M., Crainic, T.G., Nowak, M., and Rei, W. (2019). Scheduled Service Network Design with Resource Acquisition and Management under Uncertainty. *Transportation Research Part B: Methodological*, 128:324–343.

Hewitt, M., Ortmann, J., and Rei, W. (2022). Decision-based Scenario Clustering for Decision-making undewr Uncertainty. *Annals of Operations Research*, 315:747–771.

Hewitt, M., Rei, W., and Wallace, S.W. (2021). Stochastic Network Design. In Crainic, T.G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications in Transportation and Logistics*, chapter 10, pages 283–315. Springer, Boston.

Hoff, A., Lium, A.-G., Lokketangen, A., and Crainic, T.G. (2010). A Metaheuristic for Stochastic Service Network Design. *Journal of Heuristics*, 16(1):653–679.

Jarrah, A.I., Johnson, E., and Neubert, u. (2009). Large-scale, Less-than-Truckload Service Network Design. *Operations Research*, 57(3):609–625.

Jiang, X., Bai, R., W. Wallace, S., Kendall, G., and Landa-Silva, D. (2021). Soft Clustering-based Scenario Bundling for a Progressive Hedging Heuristic in Stochastic Service Network Design. *Computers and Operations Research*, 128:105182.

Kall, P. and Wallace, S.W. (1994). *Stochastic Programming*. John Wiley & Sons, Chichester, N.Y.

Kienzle, J., Crainic, T.G., Frejinger, E., and Bisaillon, S. (2024). The Intermodal Railroad Blocking and Railcar Fleet Management Planning Problem. Technical Report CIRRELT-2023, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada.

King, A.J. and Wallace, S.W.. (2012). *Modeling with Stochastic Programming*. Springer, New York.

Klibi, W., Martel, A., and Guitouni, A. (2010). The Design of Robust Value-Creating Supply Chain Networks: A Critical Review. *European Journal of Operational Research*, 203(2):283–293.

Lai, M.F. and Lo, H.K. (2004). Ferry Service Network Design: Optimal Fleet Size, Routing and Scheduling. *Transportation Research Part A: Policy and Practice*, 38:305–328.

Lange, J., Crainic, T.G., Gschwind, N., and Rei, W. (2024). An Integrated Planning Approach for a Two-tier City Logistics System with Satellite Synchronization. Publication CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Université de Montréal, Montréal, QC, Canada.

Lanza, G., Crainic, T.G., Passacantando, M., and Scutellá, M.G. (2024). A Continuous Time Physical Graph Based Formulation to Scheduled Service Network Design. In *Optimization in Green Sustainability and Ecological Transition;*, volume 2 of *AIRO Springer Series*, pages 265–275. Springer.

Lanza, G., Crainic, T.G., Rei, W., and Ricciardi, N. (2018). A Study on Travel Time Stochasticity in Service Network Design with Quality Targets. *Lecture Notes in Computer Science (LNCS)*, 11184:401–416.

Lanza, G., Crainic, T.G., Rei, W., and Ricciardi, N. (2021). Service Network Design Problem with Quality Targets and Stochastic Travel Times. *European Journal of Operational Research*, 288(1):30–46.

Lindsey, K., Erera, A., and Savelsbergh, M.W.F. (2016). Improved Integer Programming-based Neighborhood Search for Less-than-Truckload Load Plan Design. *Transportation Science*, 50(4):1360–1379.

Liu, C., Lin, S., Shen, Z.-J. M., and Zhang, J. (2023). Stochastic Service Network Design: The Value of Fixed Routes. *Transportation Research Part E: Logistics and Transportation*, 174:103118.

Lium, A.-G., Crainic, T.G., and Wallace, S.W. (2007). Correlations in Stochastic Programming: A Case from Stochastic Service Network Design. *Asia-Pacific Journal of Operational Research*, 24(2):1–13.

Lium, A.-G., Crainic, T.G., and Wallace, S.W. (2009). A Study of Demand Stochasticity in Service Network Design. *Transportation Science*, 43(2):144–157.

Macharis, C. and Bontekoning, Y.M. (2004). Opportunities for OR in Intermodal Freight Transport Research: A Review. *European Journal of Operational Research*, 153(2):400–416.

Magnanti, T.L. and Wong, R.T. (1984). Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, 18(1):1–55.

Marcucci, E., Gatta, V., and Le Pira, M., editors (2024). *Handbook on City Logistics and Urban Freight*. Edward Elgar, Northampton, MA, U.S.A.

Marshall, L., Boland, N., Savelsbergh, M.W.F., and Hewitt, M. (2021). Interval-based Dynamic Discretization Discovery for Solving the Continuous-Time Service Network Design Problem. *Transportation Science*, 55(1):29–51.

Medina, N., Hewitt, M., Lehuédé, F., and Péton, O. (2019). Integrating Long-haul and Local Transportation Planning: Service Network Design and Routing Problem. *EURO Journal on Transportation and Logistics*, 8(2):119–145.

Montreuil, B. (2011). Towards a Physical Internet: Meeting the Global Logistics Sustainability Grand Challenge. *Logistics Research*, 3(2-3):71–87.

Morganti, G., Crainic, T.G., Frejinger, E., and Ricciardi, N. (2020). Block Planning for Intermodal Rail: Methodology and Case Study. *Transportation Research Procedia*, 47:19–26.

Müller, J. P., Elbert, R., and Emde, S. (2021). Intermodal Service Network Design with Stochastic Demand and Short-term Schedule Modifications. *Computers & Industrial Engineering*, 159:107514.

Newman, A. M., Nozick, L. K., and Yano, C. A. (2002). Optimization in the rail industry. In Pardalos, P. M. and Resende, M. G. C., editors, *Handbook of Applied Optimization*, pages 704–718. Oxford University Press, New York, NY.

Ortiz-Astorquiza, ., Cordeau, J.-F., and Frejinger, E. (2021). The Locomotive Assignment Problem with Distributed Power at the Canadian National Railway Company. *Transportation Science*, 55(2):510–531.

Pedersen, M.B. and Crainic, T.G. (2007). Optimization of Intermodal Freight Service Schedules on Train Canals. Publication CIRRELT-2007-51, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Montréal, QC, Canada.

Pedersen, M.B., Crainic, T.G., and Madsen, O.B.G. (2009). Models and Tabu Search Metaheuristics for Service Network Design with Asset-Balance Requirements. *Transportation Science*, 43(2):158–177.

Piu, F. and Speranza, M.G. (2014). The Locomotive Assignment Problem: A Survey on Optimization Models. *International Transactions in Operational Research*, 21(3):327–352.

Powell, W.B. (1986a). A Local Improvement Heuristic for the Design of Less-than-Truckload Motor Carrier Networks. *Transportation Science*, 20(4):246–357.

Powell, W.B. (1986b). Approximate, Closed Form Moment Formulas for Bulk Arrival, Bulk Service Queues. *Transportation Science*, 20(1):13–23.

Powell, W.B. (2021). *Reinforcement Learning and Stochastic Optimization - A Unified Framework for Sequential Decisions*. Wiley-Interscience, Hoboken, N.J.

Powell, W.B. and Humblet, P. (1986). Queue Length and Waiting Time Transforms for Bulk Arrival, Bulk Service Queues with a General Control Strategy. *Operations Research*, 34:267–275.

Powell, W.B. and Koskosidis, Y.A. (1992). Shipment Routing Algorithms with Tree Constraints. *Transportation Science*, 26(3):230–245.

Powell, W.B. and Sheffi, Y. (1983). The Load Planning Problem of Motor Carriers: Problem Description and a Proposed Solution Approach with Predetermined Step Sizes. *Transportation Researc Part A: General*, 17(6):471–480.

Rahmaniani, R., Crainic, T.G., Gendreau, M., and Rei, W. (2017). The Benders Decomposition Algorithm: A Literature Review. *European Journal of Operational Research*, 259(3):801–817.

Ronen, D. (1983). Cargo Ships Routing and Scheduling: Survey of Models and Problems. *European Journal of Operational Research*, 12(2):119–126.

Ronen, D. (1993). Ship Scheduling: The Last Decade. *European Journal of Operational Research*, 71(3):325–333.

Roy, J. and Delorme, L. (1989). NETPLAN: A Network Optimization Model for Tactical Planning in the Less-than-Truckload Motor-Carrier Industry. *INFOR*, 27(1):22–35.

Ruszczyński, A. and Shapiro, A., editors (2003). *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. North-Holland, Amsterdam.

Scherr, Y.O., Hewitt, M., and Mattfeld, D.C. (2022). Stochastic Service Network Design for a Platooning Service Provider. *Transportation Research Part C: Emerging Technologies*, 144:103912.

Scherr, Y.O., Saavedra, B.A.N., Hewitt, M., and Mattfeld, D.C. (2019). Service Network Design with Mixed Autonomous Fleets. *Transportation Research Part E: Logistics and Transportation Review*, 124:40–55.

Smilowitz, K.R., Atamtürk, A., and Daganzo, C.F. (2003). Deferred Item and Vehicle Routing within Integrated Networks. *Transportation Research Part E: Logistics and Transportation*, 39:305–323.

SteadieSeifi, M., Dellaert, N.P. N. P., Nuijten, W., van Woensel, T., and Raoufi, R. (2014). Multimodal Freight Rransportation Planning: A Literature Review. *European Journal of Operational Research*, 233(1):1–15.

Taherkhani, G., Bilegan, I.C., Crainic, T.G., Gendreau, M., and Rei, W. (2022). Tactical Capacity Planning in an Integrated Multi-stakeholder Freight Transportation System. *OMEGA*, 110. Article 102628.

Vaidyanathan, B., Ahuja, R.K., Liu, J., and Shughart, L.A. (2008a). Real-Life Locomotive Planning: New Formulations and Computational Results. *Transportation Research Part B: Methodological*, 42(2):147–168.

Vaidyanathan, B., Ahuja, R.K., and Orlin, J.B. (2008b). The Locomotive Routing Problem. *Transportation Science*, 42(4):492–507.

Vu, D.M., Crainic, T.G., and Toulouse, M. (2013). A Three-Stage Matheuristic for the Capacitated Multi-commodity Fixed-Cost Network Design with Design-Balance Constraints. *Journal of Heuristics*, 19:757–795.

Wang, X., Crainic, T.G., and Wallace, S.W. (2019). Stochastic Network Design for Planning Scheduled Transportation Services: The Value of Deterministic Solutions. *INFORMS Journal on Computing*, 31(1):153–170.

Wieberneit, N. (2008). Service Network Design for Freight Transportation: A Review. *OR spectrum*, 30(1):77–112.

Yaghini, M. and Akhavan, R. (2012). Multicommodity Network Design Problem in Rail Freight Transportation Planning. *Procedia Social and Behavioral Sciences*, 43:728–739.

Zhu, E., Crainic, T.G.., and Gendreau, M. (2014). Scheduled Service Network Design for Freight Rail Transportations. *Operations Research*, 62(2):383–400.