CIRRELT

CIRRELT-2024-29

Continuous-Time Service Network Design with Stochastic Travel Times

Giacomo Lanza
Teodor Gabriel Crainic
Mauro Passacantando
Maria Grazia Scutellà

August 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

Continuous-Time Service Network Design with Stochastic Travel Times

Giacomo Lanza[1,*], Teodor Gabriel Crainic[2], Mauro Passacantando[3], Maria Grazia Scutellà[1]

1. Department of Computer Science, University of Pisa, Pisa, Italy
2. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and School of Management, Université du Québec à Montréal
3. Department of Business and Law, University of Milano-Bicocca, Milan, Italy

Abstract. We address the Continuous-Time Service Network Design problem with Stochastic Travel Times, to incorporate business-as-usual travel time fluctuations. We define the problem by considering quality targets for on-time operation of services and delivery of demand at destinations. We propose a two-stage mixed-integer stochastic programming model that prescribes the service selection based on probability distributions (first stage) and recourse actions using updated travel time estimations (second stage), including postponement of service departures, commodities itineraries and outsourcing, with quality targets modeled through penalties. The formulation aims to mitigate delays, reduce operational costs, and guarantee the feasibility and profitability of the solutions obtained also when travel time fluctuations occur. The formulation utilizes an innovative Stochastic-Aware Service-Leg Network to model time continuously, mitigating the increase in network size that would typically occur with traditional approaches. Through extensive experimentation on realistic small-to-medium size instances, we assess the complexity and benefits of our stochastic formulation over the deterministic one, also highlighting the specific features to hedge against time fluctuations appearing in stochastic solutions. The results demonstrate that our approach effectively mitigates delays and improves operational costs.

Keywords: Freight Transportation, Service Network Design, Stochastic Travel Time, Continuous Time, Two-Stage Formulation.

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. He gratefully acknowledges the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery grant program, and of the Fonds de recherche du Québec through their infrastructure grants.

---

* Corresponding author: giacomo.lanza@unipi.it

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

1 Introduction

Consolidation-based freight transportation carriers group and dispatch within the same vehicle or convoy (e.g., a truck, a container ship, a freight train) multiple shipments, each potentially associated with a different customer. Postal and small-package transportation companies, less-than-truckload (LTL) motor carriers, railroads, and maritime liner navigation companies perform similar services. Consolidation is essential for the profitability of these carriers as it increases resource utilization and reduces transportation costs as economies of scale can be exploited, allowing them to offer their customers competitive prices among the several competitors in this busy market.

These carriers operate within a network of terminals connected by infrastructure, such as highways or rail tracks, or conceptual links, such as maritime or air corridors. They provide transportation services along defined routes (i.e., origin, destination, and possibly intermediate stops) and according to schedules (i.e., timing information of departure from and arrival at each stop). Services may also be characterized by some physical (e.g., mode, type, capacity) and operational (e.g., priority, speed, cost) features (Crainic and Hewitt, 2021). Carriers operate services to respond to a transportation demand between specific origin and destination terminals raised by customers. Each shipment must be picked up from its origin terminal no earlier than its availability date and delivered to its destination terminal before the specified due date. Shipments are routed by services from their origins towards their destinations, through itineraries possibly visiting intermediate terminals where loading/unloading, shipment integration, and service-to-service transfer operations are performed. Typically, customers prioritize carriers' competitive pricing and value carriers' reliability in adhering to agreed-upon delivery times specified in their contracts (Crainic and Hewitt, 2021).

Jointly determining the itineraries that commodities must follow, and selecting the scheduled services to operate efficiently, profitably, and ensuring effective consolidation while maintaining the desired level of service quality that is crucial for customer satisfaction, is a rather complex problem faced at the tactical level by carriers, supported by the service network design (SND) methodology. The scope of SND is to produce a transportation plan (comprising the selected scheduled services and the commodity itineraries) which will be operated for a medium-long period with the main objective of minimizing the total transportation cost. The problem is normally addressed for a given length of time, called schedule length, for which a certain demand regularity is observed over a longer period, called a planning horizon. The transportation plan is then cyclically repeated to cover the entire planning horizon (e.g., a weekly schedule to be repeated for six months). We refer to Crainic and Hewitt (2021) for a comprehensive overview of SND problems in the context of consolidation-based transportation, their variants, and SND methodologies.

Formulations in the literature typically address the problem deterministically (the most recent contributions are discussed in Crainic and Hewitt, 2021), assuming that all necessary information to define the plan is readily available and generally fixed to specific point estimations computed through various forecasting methods from historical data. Deterministic approaches, however, often fails to adequately account for the inherent fluctuations in these problems (Powell and Topaloglu, 2003; Lium et al., 2009; Klibi et al., 2010; King and Wallace, 2012; Lanza et al., 2018; Wang et al., 2019; Hewitt et al., 2021). Indeed, like any other tactical planning problem, SND is inherently stochastic (King and Wallace, 2012; Hewitt et al., 2021). It requires carriers to make decisions in an environment with imperfect or not fully known information for parameters such as seasonal demand, travel times, transportation costs, etc., complete information becoming available only after medium-term decisions have been made. However, once established, customers and carriers expect the service network to adhere as closely as possible to the carriers' external announcements, despite potential variations in parameters - compared to the estimations used in the planning phase - observable during daily operations. The stochastic programming approach explicitly incorporates uncertainty into the decision-making process, by considering statistical distributions associated with some parameters instead of a point estimation (Birge and Louveaux, 2011). Indeed, it individuates more flexible solutions able to mitigate, or at least to limit, the adverse effects caused by fluctuations and account for the additional costs that may arise when fluctuations occur (King and Wallace, 2012; Hewitt et al., 2021).

In this paper, we investigate a stochastic Scheduled Service Network Design (SSND) problem in the context of consolidation-based freight carriers, by proposing a stochastic programming formulation that explicitly considers the stochastic nature of the *travel time*, i.e., the time required by a service to travel from one terminal to the next one along its route. Travel time uncertainty is a rarely addressed aspect in the stochastic SND literature (only a handful of works actually delve into this subject, such as Demir et al., 2016; Lanza et al., 2021; Shu et al., 2023). Our focus lies in fluctuations occurring during daily operations, the so-called randomness, i.e., foreseeable fluctuations that can be described through random variables with known probability distributions (Klibi et al., 2010). These fluctuations can have various impacts depending on the transportation modes, contexts and applications, and may be caused due to several factors, such as traffic congestion or adverse weather conditions. Depending on their severity, consequences can have ripple effects on the feasibility and profitability of the plan, leading to disruptions in the planned service-to-service transfers, hindering efficient consolidation of commodities eventually causing difficulties in meeting delivery deadlines, and ultimately, negatively impacting the reputation and revenues of the carrier.

As detailed in the next sections, the aforementioned contributions not only differ from the ones presented in this paper in terms of problem setting, formulation, and recourse actions, but also in the methodological approach used to model time within the mathematical formulation. Specifically, we propose a two-stage stochastic programming model prescribing the selection of services in the first stage, when only probability distributions for travel time information are available. Depending on the selected services and on the outcome of the travel time random variables, remaining decisions must be made. These relate to the postponement of some service departure times, to the routing of commodities through the selected services, and outsourcing. Note that these three recourse actions have never been jointly considered in the literature related to stochastic SSND. The costs associated with such decisions as well as the costs for delays related to both service operations, with respect to their schedule, and commodities arrival times at the destination, with respect to the agreed-upon time of deliveries, are included in the objective function. The goal of the proposed formulation is to design a service network that can mitigate the consequences brought by the fluctuations in travel times associated with delays in operations, jeopardizing the feasibility and profitability of the plan and generating additional costs for the carrier.

In the proposed stochastic formulation, time is modeled continuously by extending the framework in Lanza et al. (2024), which has been proposed for a deterministic SSND setting, to the stochastic setting here addressed, thus resorting to a more compact network than the traditional time-space network often used in SSND formulations. This time modeling approach complements other existing methods aimed at overcoming the issues of time-space network size growth as the level of detail in time representation increases, such the one proposed in Boland et al. (2017), where the Dynamic Discretization Discovery algorithmic framework, still reliant on time-space networks, is presented, and the ones in Hewitt and Lehuédé (2023) and Shu et al. (2023), which present formulations depending on the enumeration of consolidation opportunities. In contrast, the approach in Lanza et al. (2024) utilizes a peculiar service-leg network (where arcs describe the route of each service) that does not depend on time-space networks or the enumeration of consolidation paths.

The contributions of this paper are the following: i) we study an SSND problem setting where uncertainty in travel times is explicitly considered, with specific attention given to quality targets, and which includes recourse actions not jointly addressed by other models from the literature; ii) we propose a two-stage stochastic mixed-integer linear programming formulation for the addressed SSND with stochastic travel times; iii) we extend to the stochastic case the framework proposed in Lanza et al. (2024) for the definition of a compact network (with respect to a time-space network) on which the formulation relies, originally designed for a deterministic travel time setting; iv) we present results from extensive computational experiments on small to medium-sized realistic instances, emphasizing the advantages of our formulation in contrast to a traditional deterministic approach; in particular, the experiments highlight characteristics observed in the stochastic solutions, having the aim to reduce the adverse effects of travel time uncertainty on service network performance.

This paper is structured as follows. In Section 2, we review the sparse literature on stochastic SSND under travel time uncertainty (Section 2.1), and discuss alternative approaches to model time in SSND (Section 2.2). Section 3 describes the stochastic SSND problem addressed in this paper and introduces the mathematical notation relevant to its formulation. Section 4 is devoted to the mathematical formulation: Section 4.1 describes the stochastic-aware service-leg network on which the formulation relies, while Section 4.2 presents the two-stage stochastic mathematical model. Section 5 focuses on the computational study. Specifically, Section 5.1 describes the instances used in the study, with Section 5.1.1 discussing stability results related to the scenario generation procedure. Indeed, the computational experiments are conducted by approximating travel time probability distributions through a set of scenarios. The computational results are discussed in several sections: Section 5.2 presents numerical results evaluating the performance of the proposed stochastic model in terms of computational time and percentage optimality gap compared to a traditional deterministic approach; Section 5.3 quantifies the benefits of using the proposed stochastic approach instead of a deterministic one in terms of costs of the respective solutions; strategies for reducing delays are discussed in Section 5.4; finally, Section 5.5 analyzes how the network design configuration obtained from the stochastic model differs from that obtained from the deterministic model, using two instances as examples. Section 6 concludes the paper.

2 Literature Review

Two main contributions of this paper relate to the problem setting, which addresses stochastic travel time and recourse actions not previously jointly considered in the literature, and to the methodology proposed to the problem formulation, which relies on a more compact network than the conventional time-space often considered to model time. In the following, we therefore separately review the relevant literature on these two aspects, in Section 2.1 and in Section 2.2, respectively.

2.1 Stochastic Travel Time in Scheduled Service Network Design

Many variants of SSND in the context of freight transportation have been studied in the literature within a deterministic setting. These studies address extensions of the classical SSND, including additional management issues such as the empty repositioning of resources (e.g., vehicles or containers) and resource management considerations (Meng and Wang, 2011; Andersen et al., 2009a; Pedersen et al., 2009; Andersen et al., 2009b; Crainic et al., 2016, 2018; Hewitt et al., 2019; Scherr et al., 2019). The most recent and complete survey on SSND in specific contexts and modes of transportation, including rail, truck, and maritime, can be found in Crainic et al. (2021). Other surveys on the topic can be found in Crainic (2000); Crainic and Kim (2007); Wieberneit (2008).

SSND problems taking into account uncertainty have been studied to a lesser extent. Most of the available contributions focus on uncertainty in shipment volumes (Lium et al., 2009; Hoff et al., 2010; Bai et al., 2014; Wang et al., 2019; Wang and Qi, 2020; Müller et al., 2021b; Scherr et al., 2022). In some applications, however, the tactical plan can be defined with a demand that can be assumed to be deterministic. This is often the case in industries where demand patterns are highly stable. For example, just-in-time production systems rely on precise and predictable deliveries of raw materials to minimize inventory costs and align with the production schedule. In such systems, it is crucial to avoid delays that could halt the production line and incur significant downtime costs. Therefore, delays and consequent costs assume major importance in the planning phase, requiring it to account for variability in travel times rather than fluctuating demand.

Despite its significant importance, travel time uncertainty is very rarely addressed in the SSND literature. Among the few studies addressing such a problem, Demir et al. (2016) (and later Hrušovskỳ et al., 2018; Layeb et al., 2018) investigate an intermodal freight transportation planning problem whose goal is to route origin-destination demands by both planning road services and selecting available rail and maritime transportation services, the latter operating according to fixed schedules. The proposed stochastic mixed-integer linear programming approach incorporates probabilistic constraints ensuring that the probability of missing an available rail or maritime service for road services, due to longer than planned travel times, remains below a predefined threshold. The model does not explicitly address recourse actions or re-planning caused by missed connections at transshipment terminals. Similar intermodal SND problems have been treated in Zhao et al. (2018a,b); Sun (2020), and Müller et al. (2021a). Lanza et al. (2018, 2021) studied an SSND problem focusing on quality targets for the on-time operation of services according to a given schedule, and for the on-time delivery of commodities to destinations according to carrier-customer contracts under fluctuations in travel times. The authors propose a two-stage mixed-integer stochastic model defined over a time-space network, with quality targets modelled through penalties. The first stage of their model involves selecting services and determining freight routes. A simple recourse strategy is employed in the second stage, where penalties are paid for lateness. The problem setting considered does not include explicit penalty costs for commodities missing a transfer at a terminal due to service delays. Safe connections are only enforced through the penalties for late service arrival. Finally, in Shu et al. (2023), an SSND with uncertain travel times is addressed, focusing on delays resulting from the synchronization of multiple shipments that require consolidation at certain terminals before continuing their itineraries together. When one of these shipments is late, it can cause delays for all the others waiting it for consolidation. Furthermore, the generated delays can then be propagated and cause additional delays for other shipments at subsequent terminals. In their setting, tactical decisions involve service selection and freight routing which are not modified at an operational level, and thus no vehicle may depart until all its shipments have arrived (postponing the actual departure schedule). The setting studied in Shu et al. (2023) is the most similar to the one studied in this paper. In fact, we consider consolidation delay propagation and service departure postponement as well. However, unlike Shu et al. (2023), we propose a more flexible setting in which service selection belongs to tactical decisions and the routing of commodities - determined at an operational level - may vary across schedule repetitions based on observed travel times and delays, also including the possibility of outsourcing. Besides the problem settings, the proposed approaches are also different. In fact, Shu et al. (2023) proposes a robust formulation aimed at optimizing costs caused by delays in a worst-case scenario. In contrast, this paper follows a stochastic optimization approach to optimize costs caused by delays for average performance.

2.2 Modelling Time in Scheduled Service Network Design

In the context of SSND for tactical planning of consolidation-based freight carriers, time-related features associated with demand and services, such as availability, due dates, and schedules, are traditionally modeled using a time-space network. This approach captures the geographic and temporal components of the problem, integrating them into a single network structure. Time-space networks often resort to time discretization. Specifically, when considering a schedule length of T and selecting a granularity Δ, a time-space network is constructed by partitioning the schedule length into T/Δ non-overlapping time intervals, replicating the physical nodes $T/\Delta + 1$ times and adding arcs to connect the timed-nodes appropriately. Nodes in this network represent terminals at specific points in time, while arcs represent either service legs according to their schedule (so-called moving arcs) or the holding of goods or resources at a terminal for one period of time (so-called holding arcs). Consequently, all events taking place within a time interval are modelled as occurring at the beginning of it. Moreover, since only the scheduling period $[0, T]$ is considered in time-space networks, to have a repeatable plan the activities that would end after T, e.g., at a time $t' > T$, are modelled via arcs going back in time, i.e., arriving at their destination at time $t' \mod T$ (for a more detailed description, see Crainic and Hewitt, 2021).

The classical partitioning of the schedule length prescribes defining time intervals of equal length, by applying the same replication strategy to all physical nodes in the time-space network. Alternatively, the length of the time intervals may vary, and for some physical nodes associated with important terminals in terms of operations and workload, a finer replication is considered. Finally, the departure and arrival times of services at terminals may be used to define the relevant time instants for replicating physical nodes to be used in the time-space network, not resorting to a pure partitioning of the schedule length (Crainic, 2024).

Choosing the appropriate granularity of the schedule length partitioning can be challenging as it has an impact on both solution quality and computational tractability: the finer the granularity, the more detailed the representation of time and time-related activities, and thus, the quality of the solution. However, the replication of nodes in the time dimension can lead to a significant increase in the size of the time-space network, resulting in high computational efforts when solving the problem (see Boland et al., 2017, 2019, for an extensive discussion).

These challenges are mitigated by the Dynamic Discretization Discovery (DDD) algorithmic framework presented in Boland et al. (2017) and further studied in Hewitt (2019); Marshall et al. (2021); Shu et al. (2024). DDD is an exact algorithm that utilizes an iterative process to adjust the level of discretization of an initial and partial time-space network, without ever creating a fully time-space network in a priori manner (Shu et al., 2023). However, although DDD is computationally effective, especially on LTL applications, it can be challenging to implement, as discussed in Hewitt and Lehuédé (2023).

In the literature, recent formulations are based on more compact networks than time-space networks to mitigate the network size increase, particularly for problems requiring a detailed time representation. A consolidation-based formulation to the SSND is proposed in Hewitt and Lehuédé (2023), which is not defined on a time-space network but rather on an a priori enumeration of all or part of the possible consolidations on each arc. However, as stated by the authors, also the enumeration of the consolidation opportunities can be computationally challenging. Therefore, the authors proposed a hybrid formulation that combines elements of the consolidation-based formulation and the classical time-space network formulation. Extensions to include resource management issues, bin-packing considerations, and piecewise linear cost functions are also discussed. In the already mentioned Shu et al. (2023), a more compact *consolidation-indexed* formulation for the deterministic SSND has been proposed, where consolidation indices rather than time indices are used in defining decision variables and constraints. The formulation requires the construction of a *consolidation-expanded network*, which replicates each arc in the service-leg network according to the number of commodities in the problem, to create consolidation traveling arcs. Commodities may flow on the same arc, and thus follow the same consolidation, only if they all pass through the arc at the same time. This is ensured through specific constraints in the proposed mathematical model. The formulation has been extended to a robust version of the problem with uncertain travel times (Shu et al., 2023).

Finally, a different approach has been proposed in Lanza et al. (2024) for the deterministic SSND, relying neither on a time-space network nor on the enumeration of consolidation paths of commodities. This formulation relies on the construction of a peculiar network, called *extended service-leg network*, defined starting from the service-leg network. The extended service-leg network includes all arcs from the service-leg network, plus additional arcs modeling different occurrences of some service legs at various times (in the past or in the future). The service legs in the extended service-leg network are those that can be utilized to move the commodities from their origins to their destinations. The SSND is then formulated as a capacitated multicommodity network design problem on this network. The proposed formulation has proved to be particularly effective when a fine granularity is required, compared to traditional time-space networks, which may lead to intractable models due to the dimensional increase (Lanza et al., 2024).

The methodology outlined in Lanza et al. (2024) is well-suited for extension to the stochastic setting, where the travel time parameters associated with the service legs are replaced by probability distributions. Hereafter, by generalizing the approach in Lanza et al. (2024), the addressed stochastic SSND will be formulated over a *stochastic-aware service-leg network*, as formally defined in the following.

3 Problem Description

Let $(\mathcal{N}, \mathcal{A}^{PH})$ be the physical infrastructure network underlying the problem addressed: the set of nodes $\mathcal{N}$ represents physical terminals, while $\mathcal{A}^{PH}$ denotes the set of physical, such as highways or rail tracks, or conceptual, such as maritime or air corridors, directed links. Let T be the chosen schedule length, and thus $[0, T]$ be the scheduling period, composed of T time units. We assume each arc in $(i, j) \in \mathcal{A}^{PH}$ has an associated travel time probability distribution and a representative point estimate. These distributions account for business-as-usual travel time fluctuations, thus excluding major disturbances or catastrophic events, as described in Klibi et al. (2010). The travel time distribution for each arc (i, j) ranges between a minimum $\tau^m_{(i,j)}$ and a maximum $\tau^M_{(i,j)}$. The point estimate, termed as the *usual travel time*, is derived statistically from this distribution, representing the ideal travel time without delays (e.g., average or mode).

We represent the demand by a set of commodities $\mathcal{K}$. Each commodity $k \in \mathcal{K}$ requires the transport of a certain volume d^k from an origin terminal $O(k) \in \mathcal{N}$ to a destination terminal $D(k) \in \mathcal{N}$, according to its availability date $o(k) \in [0, T]$ at the origin and a given due date $d(k)$ at the destination. Demand features are assumed not to be affected by random fluctuations in this setting. Furthermore, we assume that each commodity must follow a single directed path from its origin to its destination.

The set of potential services the carrier may operate to meet transportation demand, which we assume move all at the same speed, is denoted by Σ. Each potential service

$\sigma \in \Sigma$ is defined by its *route* in the physical network and by its schedule. The route of potential service σ is defined through the set of its service legs $\mathcal{L}(\sigma)$. If there are no intermediate stops, $\mathcal{L}(\sigma)$ includes a single leg, i.e., $\mathcal{L}(\sigma) = \{l(\sigma)\}$, from the origin $O(\sigma)$ to the destination $D(\sigma)$. If there are intermediate stops, $\mathcal{L}(\sigma)$ contains multiple legs, i.e., $\mathcal{L}(\sigma) = \{l_i(\sigma) : \; i = 1, \dots, |\mathcal{L}(\sigma)|\}$, each connecting consecutive terminals on the route. Note that each service leg defines a directed path, in the physical network, connecting two consecutive terminals on the route of a service.

At this regard, we introduce the *service-leg* network $(\mathcal{N}, \mathcal{A})$, where $\mathcal{A} = \bigcup_{\sigma \in \Sigma} \mathcal{L}(\sigma)$ is the set of the service legs of all the potential services in Σ. Note that $\mathcal{A}$ may contain multiple parallel arcs, i.e., arcs having the same origin and destination, being however associated with different services. We define $\sigma_a \in \Sigma$ as the service associated with arc $a \in \mathcal{A}$, and u_a as the capacity available on service leg a.

For each service leg $a \in \mathcal{A}$, we define the travel time random variable τ_a as the convolution of the travel times random variables of the physical arcs composing the service leg, and we denote by $[\tau_a^m, \tau_a^M]$ the finite support of τ_a. Moreover, $\hat{\phi}_a$ denotes the usual time instant at which service σ_a begins to execute (i.e., to move along) arc a, while $\hat{\psi}_a$ denotes the usual time instant at which service σ_a ends to execute (i.e., to move along) arc a. The latter is built considering the usual travel times. Finally, for those services including intermediate stops, we define $\theta_{l_i(\sigma)}$ the time that service σ must spend at the terminal being the head of service leg $l_i(\sigma)$, with $i = 1, \dots, |\mathcal{L}(\sigma)| - 1$. The latter parameter is assumed to be deterministic. Thus, the *usual schedule* of a service is defined considering the usual inter-terminal travel times and the deterministic operation time at the visited intermediate terminals.

In this setting, we consider the following information revelation and decision-making process. Service selection is made at the planning stage, before observing the actual travel time realizations, thus only taking into account their probability distributions. Once these decisions are made, and just before service operations begin, i.e., at the beginning of the schedule period, we assume we can obtain a more precise estimation of travel times based on system conditions. Improving the accuracy of travel time estimations shortly before transportation services depart is not only possible but also highly advantageous, particularly considering the dynamic nature of factors such as traffic and weather conditions. The new information refining travel time estimations is then used to make additional decisions regarding service operations and commodity movements. We determine the actual departure times of each selected service from the visited terminals on its route, within operational feasibility limits. More precisely, regarding service operations, when the travel time on an arc a is shorter than the usual time, then service σ_a arrives earlier at the head terminal and must wait until the scheduled time, $\hat{\psi}_a$, to proceed (to respect the strict schedule terminals have as well). Conversely, if service σ_a arrives later due to a longer travel time, terminal operations begin immediately, and the departure time is adjusted to account for the late arrival, propagating the delay to the next leg of the route (recall that terminal operations take a deterministic amount of time). Additionally, delays can result from synchronizing multiple commodities at a terminal needing consolidation on the same service. If one of those is late, the departure time of a service may be adjusted, i.e., postponed, to ensure smooth consolidation. Specifically, the departure of service σ_a may be postponed by at most δ_a units of time. Such postponements can cause tardy deliveries. In this respect, we also assume that each commodity k may arrive at its destination with a delay of at most δ^k after its delivery time. Furthermore, the information refining travel time estimations is also used to establish the actual itinerary commodities follow through the selected services. The delivery of a commodity can also be outsourced, in which case its entire volume is moved from the origin to the destination by an expensive external service provider.

Each of the decisions described above causes some costs. Specifically, we define f_σ as the activation cost for service $\sigma \in \Sigma$. Moreover, c_a^k represents the unit commodity transportation cost for commodity $k \in \mathcal{K}$ and arc $a \in \mathcal{A}$, while c_i^k denotes the unitary cost of holding or handling goods of type $k \in \mathcal{K}$ at node $i \in \mathcal{N}$. Delayed service arrivals and commodity deliveries incur costs, too: q_a is the fixed cost to pay if $\sigma_a \in \Sigma$ finishes executing the movement on arc a beyond the usual arrival time $\hat{\psi}_a$, while q^k is the unitary cost to pay if commodity $k \in \mathcal{K}$ arrives at its destination beyond the due date $d(k)$. Notice that delays in service arrivals at terminals do not always imply late deliveries of commodities. For instance, subsequent legs with shorter travel times can often accommodate delays without disrupting commodity itineraries. Conversely, commodities may arrive late at their destinations even if services adhere to their schedules. Finally, C^k represents the unitary cost incurred when outsourcing commodity $k \in \mathcal{K}$. The notation introduced in this section is summarized in Table 10.

4 Mathematical Formulation

In this section, we first describe the stochastic-aware service-leg network which allows us to model time continuously, Then, we present the two-stage stochastic mathematical model based on it.

4.1 Stochastic-Aware Service-Leg Network

We rely on the following assumptions regarding the set of potential services Σ:

$$\hat{\phi}_{l_1(\sigma)} \in [0, T] \qquad \forall\, \sigma \in \Sigma, \tag{1}$$

$$\hat{\phi}_{l_{|\mathcal{L}(\sigma)|}(\sigma)} + \delta_{l_{|\mathcal{L}(\sigma)|}(\sigma)} + \tau^M_{l_{|\mathcal{L}(\sigma)|}(\sigma)} < 2T \qquad \forall\, \sigma \in \Sigma. \tag{2}$$

Conditions (1) ensure that the usual departure time of each service σ occurs within the scheduling period $[0, T]$, while conditions (2) impose that the arrival time of each service at its destination, also in the worst case allowed by the stochastic setting under consideration, occurs before $2T$. Notice that the arrival time at destination of a service σ in the worst case is given by the sum of the usual departure time of σ to execute the last leg on its route, i.e., $\hat{\phi}_{l_{|\mathcal{L}(\sigma)|}(\sigma)}$, the maximum permissible postponement of the departure, i.e., $\delta_{l_{|\mathcal{L}(\sigma)|}(\sigma)}$, and the maximum observable travel time of such a last leg, i.e., $\tau^M_{l_{|\mathcal{L}(\sigma)|}(\sigma)}$.

Moreover, regarding demand, we assume the following:

$$o(k) \in [0, T] \qquad \forall\, k \in \mathcal{K}, \tag{3}$$

$$d(k) + \delta^k \leq o(k) + T \qquad \forall\, k \in \mathcal{K}. \tag{4}$$

Conditions (3) ensure that the availability date of each commodity falls within the scheduling period $[0, T]$, while conditions (4) ensure that the duration of the itinerary of each commodity from its origin to its destination is at most T, also in case of delay.

As mentioned, the proposed mathematical model is based on a stochastic-aware service-leg network $(\mathcal{N}, \mathcal{A}^{STT})$. Such a network accounts for both the repetitiveness of the tactical plan answering the regularity in demand over the planning horizon and the fact that not all the demand and services have their associated time attributes within the schedule length. In fact, some may occur during the previous application of the plan and terminate in the current one, others while may start during the current application, may terminate during the following one.

Informally speaking, the set of arcs $\mathcal{A}^{STT}$ includes all those service legs that can be exploited by the commodities in $\mathcal{K}$ to reach their destinations, while accounting for possible delays. In fact, $\mathcal{A}^{STT}$ contains transpositions in time of some service legs in $\mathcal{A}$. That is to say, considering the service legs of $\mathcal{A}$, in $\mathcal{A}^{STT}$ additional occurrence T time units before or T time units ahead with respect to their associated usual schedule time are included.

More formally, let us define the following two subsets of $\mathcal{A}$:

$$\mathcal{A}^+ := \{a \in \mathcal{A} : \; \hat{\psi}_a \leq T\} \qquad \text{and} \qquad \mathcal{A}^- := \{a \in \mathcal{A} : \; \hat{\phi}_a + \delta_a \geq T\}. \tag{5}$$

$\mathcal{A}^+$ includes those service legs in $\mathcal{A}$ representing movements ending before T or exactly at T according to their usual schedule, while $\mathcal{A}^-$ includes those service legs in $\mathcal{A}$ representing movements beginning after T or exactly at T when considering their maximum possible leaving delay. Notice that, according to (5), $\mathcal{A}^-$ includes those service legs with usual departure time less than T, which however might leave after T due to a departure postponement.

Definition 1 *The set of arcs $\mathcal{A}^{STT}$ of the stochastic-aware service-leg network contains:*

- *all the arcs in $\mathcal{A}$;*
- *for each $a \in \mathcal{A}^-$, one additional arc, named a^-, with the same head and tail of a, representing the previous occurrence of activity a performed T time units in the past;*
- *for each $a \in \mathcal{A}^+$, one additional arc, named a^+, with the same head and tail of a, representing the next occurrence of activity a performed T time units ahead.*

Therefore:

$$\begin{aligned}
&\hat{\phi}_{a^+} = \hat{\phi}_a + T \quad \forall\, a \in \mathcal{A}^+, && \hat{\phi}_{a^-} = \hat{\phi}_a - T \quad \forall\, a \in \mathcal{A}^-,\\
&\hat{\psi}_{a^+} = \hat{\psi}_a + T \quad \forall\, a \in \mathcal{A}^+, && \hat{\psi}_{a^-} = \hat{\psi}_a - T \quad \forall\, a \in \mathcal{A}^-,\\
&c^k_{a^+} = c^k_a \quad \forall\, a \in \mathcal{A}^+, && c^k_{a^-} = c^k_a \quad \forall\, a \in \mathcal{A}^-,\\
&q_{a^+} = q_a \quad \forall\, a \in \mathcal{A}^+, && q_{a^-} = q_a \quad \forall\, a \in \mathcal{A}^-,\\
&\delta_{a^+} = \delta_a \quad \forall\, a \in \mathcal{A}^+, && \delta_{a^-} = \delta_a \quad \forall\, a \in \mathcal{A}^-.
\end{aligned} \tag{6}$$

The rationale underlying Definition 1 is the following: since, according to (3), the availability of each commodity is in $[0,\, T]$, and, according to (4), the commodity itinerary lasts at most T, then the time interval comprised between 0 and $2T$ must indeed to be considered to plan the commodity itinerary. Consequently, for each service leg in $\mathcal{A}$ whose usual ending time is less than or equal to T (defining subset $\mathcal{A}^+$), then its occurrence starting T time units ahead could be used to move some commodities and therefore has to be considered in planning the commodity itinerary. Similarly, for each service leg in $\mathcal{A}$ whose starting time is greater than or equal to T in the worst situation allowed by the addressed stochastic framework, i.e., in the case of maximum possible delay (defining subset $\mathcal{A}^-$), then its occurrence starting T time units before could be used to move some commodities, and therefore has to be considered in planning the commodity itinerary. Notice, in particular, that if a belongs to $\mathcal{A}^-$ and it departs before T according to its

Table 1: Data of Example 1.

(a) Commodity requirement.

$\mathcal{K}$	$O(k)$	$D(k)$	$o(k)$	$d(k)$	d^k	δ^k
k_1	A	C	1	8	1	0
k_2	B	A	9	19	1	0
k_3	A	B	8	15	1	0
k_4	A	C	0	4	1	1

(b) Route and schedule of services.

$\mathcal{A}$	$\hat{\phi}_a$	$\hat{\psi}_a$	θ_a	u_a	δ_a
$l_1(\sigma):\ C \to A$	6	9	2	4	0
$l_2(\sigma):\ A \to B$	11	12	0	4	0
$l(\xi):\ B \to C$	3	5	0	4	0
$l(\zeta):\ B \to C$	9	11	0	4	2

usual schedule, then the departure time of a^- is a negative value since it is shifted by $-T$ (this is exemplified in Example 1). Note also that a service leg a starting before T and ending after T does not belong to either $\mathcal{A}^+$ or $\mathcal{A}^-$, and therefore no additional arc associated with a is included in the set $\mathcal{A}^{STT}$. In fact, the additional arcs a^- and a^+ would correspond to service legs that no commodity can utilize for its transportation since the availability of each commodity is in $[0, T]$ and, due to (4), its itinerary may last at most T.

We illustrate the definition of the stochastic-aware service-leg network through the following example.

Example 1 *Consider a schedule length $T = 10$, a scheduling period $[0, 10]$, a network composed of three terminals, $\mathcal{N} = \{A, B, C\}$, and a transportation demand composed of four commodities, $\mathcal{K} = \{k_1, k_2, k_3, k_4\}$. Commodity k_1 needs to be moved from A to C, k_2 from B to A, k_3 from A to B, and k_4 from B to C. Each commodity has a given availability, due date and volume, shown in Table 1a. Also, suppose that three scheduled potential services are available for commodity transportation, $\Sigma = \{\sigma, \xi, \zeta\}$, where $\mathcal{L}(\sigma) = \{l_i(\sigma) :\ i = 1, 2\}$, $\mathcal{L}(\xi) = \{l(\xi)\}$, and $\mathcal{L}(\zeta) = \{l(\zeta)\}$. More precisely, the route of potential service σ consists of two service legs, $l_1(\sigma)$ defining the leg from C to the intermediate stop A and $l_2(\sigma)$ defining the leg from A to B, while the route of both potential services ξ and ζ consists of a single service leg from B to C. The schedule and the capacity of the potential services are reported in Table 1b (recall that θ denotes the idle time of a service at a terminal).*

In this example, we assume that the maximum allowed delay for commodity k_4 at its destination is 1, i.e., $\delta^{k_4} = 1$, the maximum allowed departure postponement from the origin for service ζ is 2, i.e., $\delta_{l(\zeta)} = 2$, and all the other possible delays are 0.

Figure 1 provides a graphical representation of the data in Table 1 through a time chart. The horizontal axis represents time, starting from time instant -1 *to time instant 20. The vertical axis represents geographical locations, i.e., the three terminals A, B and C. Events indicating the usual starting time and the usual ending time of service legs are depicted with black round markers at the corresponding time instant and geographical location. Each arrow represents a service leg, with head and tail marking the start and the end locations and the related usual times. Events representing the availability and due dates of commodities are also shown, denoted by solid arrows (i.e., filled with a different colour for each associated commodity) pointing downwards or upwards, respectively. Note that, according to conditions* (3)*, the availability dates of all the commodities fall within the scheduling period* $[0, 10]$.

The possible maximum departure postponement of service ζ *in executing service leg* $l(\zeta)$ *is highlighted by a dashed black line labelled* $\delta_{l(\zeta)}$*, extending from the usual departure time to the maximum allowable postponement time. Possible outcomes of the travel time for service* ζ *along its service leg are depicted by dashed (brown) arrows. The varying lengths of the arcs correspond to the different service times along those routes. The plain arrow represents the travel time without perturbation. Moreover, the maximum allowed delay at the destination for commodity* k_4 *is displayed as well, and it is denoted by* δ^{k_4}.

The set of service legs for the considered example is therefore $\mathcal{A} = \{l_1(\sigma), l_2(\sigma), l(\xi), l(\zeta)\}$*, that is, the service legs depicted in Figure 1. The set* $\mathcal{A}$ *includes the subsets* $\mathcal{A}^+ = \{l_1(\sigma), l(\xi)\}$ *and* $\mathcal{A}^- = \{l_2(\sigma), l(\zeta)\}$*. Hence, the set of arcs* $\mathcal{A}^{STT}$ *of the corresponding stochastic-aware service-leg network includes:*

- *all the service legs in* $\mathcal{A}$*, i.e.,* $\{l_1(\sigma), l_2(\sigma), l(\xi), l(\zeta)\}$*;*

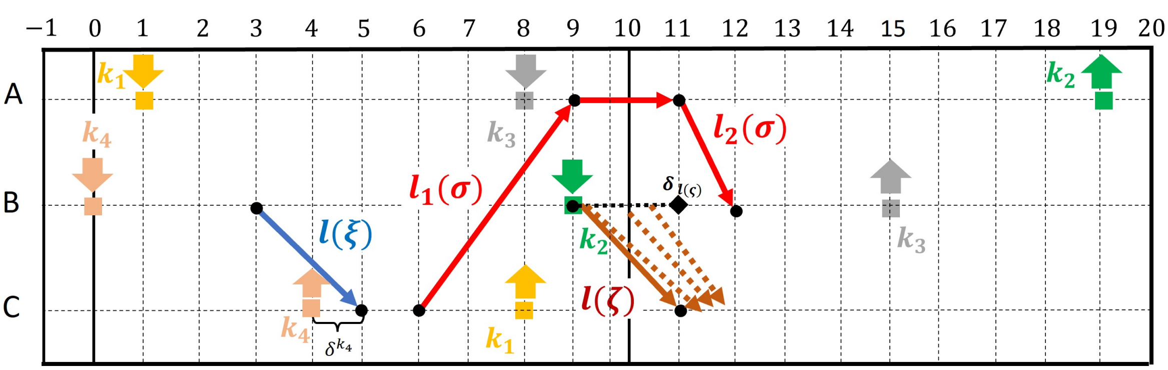


Figure 1: Set $\mathcal{A}$ associated with Example 1.

- *two additional service legs $l_1^+(\sigma)$ and $l^+(\xi)$, whose features are defined by the first set of conditions* (6)*;*

- *two additional service legs $l_2^-(\sigma)$ and $l^-(\zeta)$, whose features are defined by the second set of conditions* (6)*.*

Consequently, $\mathcal{A}^{STT} = \{l_1(\sigma),\, l_2(\sigma),\, l(\xi),\, l(\zeta),\, l_1^+(\sigma),\, l^+(\xi),\, l_2^-(\sigma),\, l^-(\zeta)\}$. Figure 2 shows the set $\mathcal{A}^{STT}$ on a time chart with the same features as the one in Figure 1. The figure clearly shows that the service legs in $\mathcal{A}^{STT}$ are those that the four commodities may exploit to reach their destination.

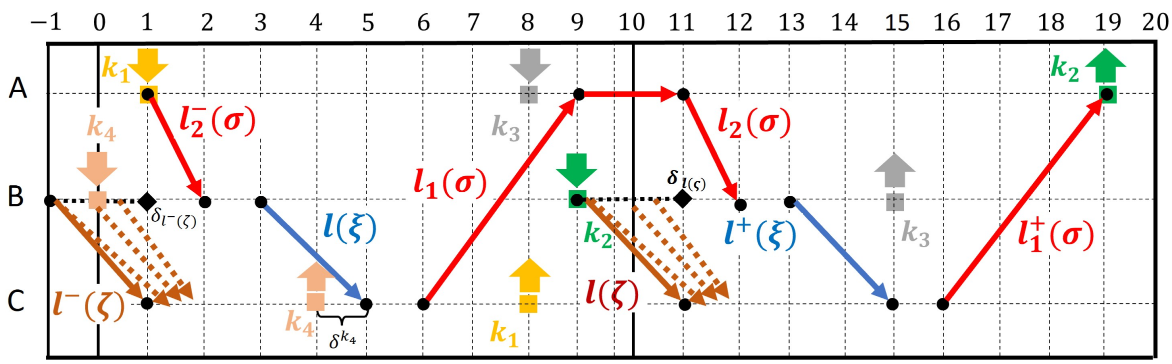


Figure 2: Set $\mathcal{A}^{STT}$ associated with Example 1.

The mathematical model we propose determines the transportation plan, for commodities having the availability date within the scheduling period $[0, T]$ and an itinerary lasting at most T, as assumed, by formulating the considered stochastic SSND in terms of a capacitated multicommodity network flow problem over the stochastic-aware service-leg network $(\mathcal{N}, \mathcal{A}^{STT})$. Such a model includes flow conservation constraints, linking-capacity constraints as well as service and commodities time management constraints, as deeply described in Section 4.2. Figure 3 shows the stochastic-aware service-leg network $(\mathcal{N}, \mathcal{A}^{STT})$ associated with Example 1.

Notice the presence of parallel arcs in such a network, that may represent either the operation of a service leg of the same service in different time periods (e.g., $l(\zeta)$, $l^-(\zeta)$) (in such a case, the different timing parameters associated with the parallel arcs differentiate the alternatives), or the operation of service legs of different services in the same scheduling period (e.g., $l(\zeta)$, $l(\xi)$).

The proposed formulation includes a sort of bundle variables associated with service activation. In fact, if a service is activated, then all its service legs are available for

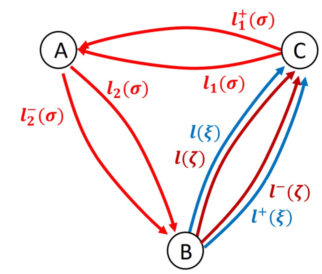


Figure 3: Stochastic-aware service-leg network associated with Example 1.

transportation, considering also the possible transposition in time of them according to the associated stochastic-aware service-leg network. Moreover, the formulation ensures that each service has enough capacity to transport the assigned commodities, through the linking-capacity constraints, by imposing that the sum of the volumes of the commodities transported on the arcs associated with the same service leg (i.e., on a^+ or a^- in case of a service leg a) does not exceed the capacity of the service. This concept is better illustrated through the continuation of the Example 1.

Example 1 (continued) *A possible solution to the considered instance is depicted in Figure 4, where the itineraries of the four commodities* k_1*,* k_2 k_3 *and* k_4 *are outlined using yellow, green, grey, and orange arrows, respectively.*

The figure shows that both k_1 *and* k_3 *use the second leg of service* σ*. Specifically, by considering the stochastic-aware service-leg network,* k_1 *is moved on* $l_2^-(\sigma)$ *while* k_3 *is moved on* $l_2(\sigma)$*. Since* $l_2^-(\sigma)$ *and* $l_2(\sigma)$ *indeed refer to the same service* σ*, with* $l_2^-(\sigma)$ *representing the same transportation segment modeled by* $l_2(\sigma)$*, but performed* $T = 10$ *time units before, and given the regularity in transportation demand, i.e., its repetition every* $T = 10$ *time units, we need to impose that the sum of the volumes of the commodities*

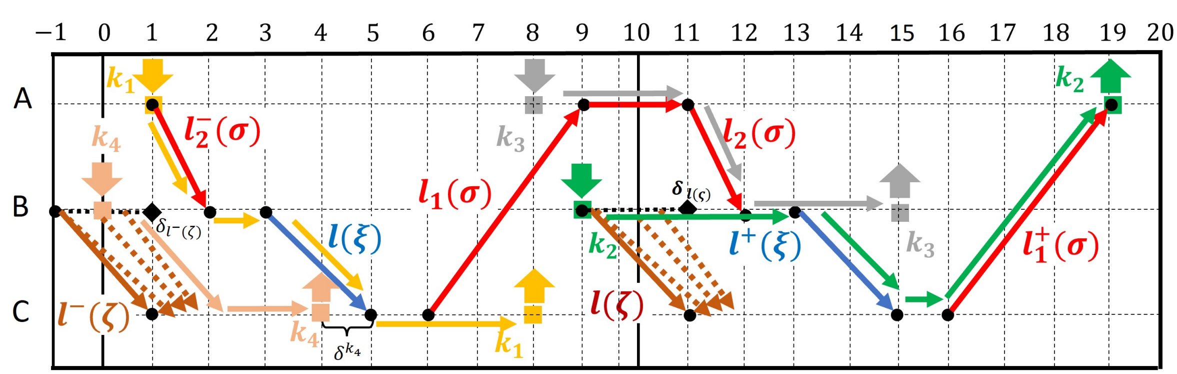


Figure 4: A possible solution to Example 1.

k_1 and k_3 does not exceed the capacity of the service leg $l_2(\sigma)$. Looking at the itineraries of k_1 and k_2, the same considerations can be made for service legs $l(\xi)$ and $l^+(\xi)$. Finally, observe that there are two options available for transporting commodity k_4. The first option entails utilizing service leg $l(\xi)$, which transports k_4 to its destination after its due date, albeit still within the maximum allowable delay for the delivery of commodity k_4. On the other hand, the second option involves postponing the departure of service leg $l^-(\zeta)$ to a time instant comprised between 0 and 1, as shown in the figure, thus enabling k_4 to reach its destination before the due date.

The proposed mathematical model on the stochastic-aware service-leg network will be presented in the next section.

4.2 Two-stage Stochastic SSND Model

As mentioned, the initial decision-making stage pertains to the tactical selection of services, based solely on known statistical distributions of travel times. This constitutes the first stage of the proposed stochastic model, where service selection is addressed. Subsequently, once those decisions are made, additional information becomes available before the execution of the service operations, enhancing the accuracy of estimated travel times across the entire network. Utilizing these enhanced insights, recourse actions can be undertaken, including: (i) postponement of service departures on service legs, (ii) definition of itineraries for commodities in alignment with the selected services, and (iii) potential outsourcing of some commodities from their origin to their destination. Moreover, the costs to pay for late arrival of services at their stops with respect to the usual arrival times, and the late arrival of commodities at their destination with respect to the agreed-upon time of deliveries, can be computed.

Given the information revelation and the decision-making assumed, the problem can be formulated as a two-stage stochastic programming model encompassing planning and recourse phases. The goal is to minimize the overall cost of the system across the two phases, i.e., the costs associated with activating services at the tactical level (first stage), as well as the average costs at the operational level depending on service selection and travel time distribution outcomes (second stage), namely commodity routing and holding costs, cost of outsourcing and costs related to delays of services at stops and of commodities at their destination. At this regard, we emphasize that the formulation proposed in Shu et al. (2023) is instead a robust formulation, aiming at optimizing costs in a worst-case scenario, besides the additional differences in problem setting already outlined in Section 3.

As mentioned, the uncertain travel time related to a service leg $a \in \mathcal{A}$ is represented by a random variable τ_a, which takes values within a finite support between a minimum τ_a^m and a maximum τ_a^M on a probability space Ω. The latter represents the entire set of possible outcomes of the random variable τ_a related to a service leg $a \in \mathcal{A}$. Let ω denote an element of this set, namely, a possible outcome in the probability space Ω. The realized travel time, given a specific outcome ω, is denoted by $\tau_a(\omega)$.

To formulate the problem we define nine sets of variables. The first stage variables are:

- $y_\sigma \in \{0, 1\}$, $\sigma \in \Sigma$, indicates whether service σ is selected ($y_\sigma = 1$), or not ($y_\sigma = 0$).

The second stage variables are:

- $x_a^k(\omega) \in \{0, 1\}$, $k \in \mathcal{K}$, $a \in \mathcal{A}^{STT}$, indicates whether commodity k moves along arc a;
- $\pi_a(\omega) \in [0,\, \delta_a]$, $a \in \mathcal{A}^{STT}$, represents the postponement relative to the usual time instant at which service σ_a begins to execute (move along) arc a;
- $\psi_a(\omega) \geq 0$, $a \in \mathcal{A}^{STT}$, represents the time instant at which service σ_a ends to execute (move along) arc a;
- $\varepsilon_i^k(\omega) \geq 0$, $k \in \mathcal{K}$, $i \in \mathcal{N} \setminus \{D(k)\}$, represents the time instant at which commodity k begins its movement from terminal i;
- $\eta_i^k(\omega) \geq 0$, $k \in \mathcal{K}$, $i \in \mathcal{N} \setminus \{O(k)\}$, represents the time instant at which commodity k ends its movement to terminal i;
- $r_a(\omega) \geq 0$, $a \in \mathcal{A}$, defines the tardiness in ending to execute (move along) arc a with respect to the usual ending time;
- $r^k(\omega) \geq 0$, $k \in \mathcal{K}$, defines the tardiness with respect to the usual due date of commodity k at its destination;
- $z^k(\omega) \in \{0, 1\}$, $k \in \mathcal{K}$, indicates whether commodity k is outsourced from $O(k)$ to $D(k)$.

Mathematical formulation. The problem is formulated as follows:

$$\min \sum_{\sigma\in\Sigma} f_\sigma y_\sigma + \mathbb{E}_{\tau_a, a\in\mathcal{A}^{STT}}[Q(y;\tau_a(\omega))] \tag{7}$$

where $Q(y;\tau_a(\omega))$ is defined as follows:

$$\begin{aligned} Q(y;\tau_a(\omega)) = \min \Bigg\{ d^k \Bigg[& \sum_{k\in\mathcal{K}} \sum_{a\in\mathcal{A}^{STT}} c_a^k x_a^k(\omega) \\ & + \sum_{k\in\mathcal{K}} \sum_{\substack{i\in\mathcal{N}:\\ i\neq O(k),D(k)}} c_i^k \left(\varepsilon_i^k(\omega) - \eta_i^k(\omega)\right) \\ & + \sum_{k\in\mathcal{K}} c_{O(k)}^k \left(\varepsilon_{O(k)}^k(\omega) - o(k)\right) + \sum_{k\in\mathcal{K}} c_{D(k)}^k \left(d(k) - \eta_{D(k)}^k(\omega) + r^k(\omega)\right) \\ & + \sum_{k\in\mathcal{K}} q^k r^k(\omega) + \sum_{k\in\mathcal{K}} C^k z^k(\omega) \Bigg] \\ & + \sum_{a\in\mathcal{A}} q_a r_a(\omega) \Bigg\} \end{aligned} \tag{8}$$

subject to the following constraints:

Commodity routing management

$$\sum_{a\in\mathcal{A}^{STT}:D(a)=i} x_a^k(\omega) - \sum_{a\in\mathcal{A}^{STT}:O(a)=i} x_a^k(\omega) = \begin{cases} -1 + z^k(\omega) & \text{if } i = O(k), \\ 1 - z^k(\omega) & \text{if } i = D(k), \\ 0 & \text{otherwise} \end{cases} \quad \begin{matrix} \forall\, k\in\mathcal{K}, \\ \forall\, i\in\mathcal{N}, \end{matrix} \tag{9}$$

$$\sum_{k\in\mathcal{K}} d^k \left[x_a^k(\omega) + x_{a^+}^k(\omega)\right] \le u_a\, y_{\sigma_a} \qquad \forall\, a\in\mathcal{A}^+ \tag{10}$$

$$\sum_{k\in\mathcal{K}} d^k \left[x_{a^-}^k(\omega) + x_a^k(\omega)\right] \le u_a\, y_{\sigma_a} \qquad \forall\, a\in\mathcal{A}^- \tag{11}$$

$$\sum_{k\in\mathcal{K}} d^k\, x_a^k(\omega) \le u_a\, y_{\sigma_a} \qquad \forall\, a\in\mathcal{A}\setminus(\mathcal{A}^+\cup\mathcal{A}^-) \tag{12}$$

Service time management

$$\hat{\phi}_{l_{i+1}(\sigma)} + \pi_{l_{i+1}(\sigma)}(\omega) \ge \psi_{l_i(\sigma)}(\omega) + \theta_{l_i(\sigma)}\, y_\sigma \qquad \forall\,\sigma\in\Sigma,\ i = 1,\dots,|\mathcal{L}(\sigma)|-1 \tag{13}$$

$$\psi_a(\omega) = \hat{\phi}_a + \pi_a(\omega) + \tau_a(\omega)\, y_{\sigma_a} \qquad \forall\, a\in\mathcal{A} \tag{14}$$

$$\psi_a(\omega) \le \hat{\psi}_a + r_a(\omega) \qquad \forall\, a\in\mathcal{A} \tag{15}$$

$$\pi_{a^+}(\omega) = \pi_a(\omega) \qquad \forall\, a \in \mathcal{A}^+ \tag{16}$$

$$\psi_{a^+}(\omega) = \psi_a(\omega) + T \qquad \forall\, a \in \mathcal{A}^+ \tag{17}$$

$$\pi_{a^-}(\omega) = \pi_a(\omega) \qquad \forall\, a \in \mathcal{A}^- \tag{18}$$

$$\psi_{a^-}(\omega) = \psi_a(\omega) - T \qquad \forall\, a \in \mathcal{A}^- \tag{19}$$

Commodity time management

$$\varepsilon^k_{O(k)}(\omega) \geq o(k) \qquad \forall\, k \in \mathcal{K} \tag{20}$$

$$\varepsilon^k_i(\omega) \geq \eta^k_i(\omega) \qquad \forall\, k \in \mathcal{K},\ \forall\, i \in \mathcal{N} \setminus \{O(k), D(k)\} \tag{21}$$

$$\varepsilon^k_i(\omega) \leq 2T \sum_{a \in \mathcal{A}^{STT}:\, O(a)=i} x^k_a(\omega) \qquad \forall\, k \in \mathcal{K},\ \forall\, i \in \mathcal{N} \setminus \{O(k), D(k)\} \tag{22}$$

$$\varepsilon^k_{O(a)}(\omega) \geq \hat{\phi}_a + \pi_a(\omega) - 2T\left(1 - x^k_a(\omega)\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT} \tag{23}$$

$$\varepsilon^k_{O(a)}(\omega) \leq \hat{\phi}_a + \pi_a(\omega) + 2T\left(1 - x^k_a(\omega)\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT} \tag{24}$$

$$\eta^k_{D(a)}(\omega) \geq \psi_a(\omega) - 2T\left(1 - x^k_a(\omega)\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT} \tag{25}$$

$$\eta^k_{D(a)}(\omega) \leq \psi_a(\omega) + 2T\left(1 - x^k_a(\omega)\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT} \tag{26}$$

$$r^k(\omega) \geq \eta^k_{D(k)}(\omega) - d(k) \qquad \forall\, k \in \mathcal{K} \tag{27}$$

$$r^k(\omega) \leq \delta^k \qquad \forall\, k \in \mathcal{K} \tag{28}$$

The objective function (7), to be minimized, represents the cost of operating the transportation network. It includes the costs associated with the selection of the services (first term) plus the expectation, over the travel time probability distributions, of the minimum costs to be paid for the given service selection. The latter is specified in (8) and it is given by the routing costs of commodities (first term), holding and handling costs of commodities at intermediate stops (second term), at their origin (third term), and at their destination (fourth term). Moreover, it includes the costs for commodities arriving late at their destinations compared to the usual due dates and the cost of outsourcing commodities (fifth and sixth terms, respectively). Finally, it comprises the costs for services arriving late at one of the stops along their route compared to the usual schedules (seventh term).

Constraints (9) are not conventional flow conservation constraints. They ensure that each commodity k is routed from its origin to its destination through a single path when the variable $z^k(\omega)$ equals 0, i.e., the commodity is not outsourced. In that case, a feasible path must be determined for commodity k from its origin to its destination, given the activated services and the outcomes of the travel time random variables. As opposed, when variable $z^k(\omega)$ equals 1, a path is not determined for commodity k, which is thus routed at the destination through outsourcing. Note that, in such a case, all the variables $x_a^k(\omega)$ equal 0 for commodity k at the optimal solution level due to cost minimization. Finally, notice also that subtours are excluded at the optimal solution level, again for cost minimization. Inequalities (10)–(12) describe linking-capacity constraints. They ensure that commodities can be routed on a service leg a only if the service σ_a operating on that leg is selected for activation. Additionally, these relations impose a limit on the total quantity of commodities that can be routed on service legs. Specifically, for each service leg $a \in \mathcal{A}^+$, constraints (10) ensure that the total volume of demand flowing on arcs a and a^+ does not exceed the capacity available on service leg a (recall that a^+ represents the same operations performed for arc a, but translated T times ahead); similarly, for each service leg $a \in \mathcal{A}^-$, constraints (11) enforce this limitation for arcs a and a^- (recall that a^- represents the same operations performed for arc a, but translated T times in the past); lastly, for each service $a \in \mathcal{A} \setminus (\mathcal{A}^+ \cup \mathcal{A}^-)$, constraints (12) guarantee that the volume of demand flowing on arc a is within the capacity available on service leg a.

Constraints (13)–(19) relate to the time management of services. Constraints (13) ensure that, for those services σ including an intermediate stop, the time instant at which the service begins executing its service leg $l_{i+1}(\sigma)$ on its route cannot be earlier than the time instant at which the service has ended executing its previous leg $l_i(\sigma)$, plus the operation time $\theta_{l_i(\sigma)}$ required at the end of the $l_i(\sigma)$ execution. Constraints (14) define the time instant at which service σ_a ends executing arc a as the time instant at which the service σ_a begins executing arc a plus the travel time on arc a. Note that such a time instant depends on the realization of the travel time random variable $\tau_a(\omega)$ associated with arc a. Through constraints (15) the tardiness of arc a is defined, calculated as the difference between the time instant at which the movement usually ends and the time instant at which it ended (depending on the travel time random variable realization on arc a). Constraints (16)–(19) tie the time management of arc a, operated by service σ_a, with its past (16)–(17) or future (18)–(19) occurrences, if any.

Finally, constraints (20)–(28) relate to the time management for commodities. Constraints (20) ensure that each commodity departs from its origin not before it becomes available. For those nodes i being neither the origin nor the destination of a commodity k, constraints (21) guarantee that the leaving time of k from i be greater than or equal to the time at which k arrived at that node; as opposed, if commodity k does not pass through terminal i, then constraints (22) force to 0 variables $\varepsilon_i^k(\omega)$. Constraints (23)–(24) ensure that the time instant $\epsilon_{O(a)}^k$ at which a commodity k begins executing arc a equals the time instant at which the service associated with arc a begins executing it, if k is loaded on that service. More precisely, since T represents the length of the scheduling period and any commodity may only follow a single path from its origin to its destination, when commodity k is moved along a, i.e., $x_a^k(\omega) = 1$, then constraints (23)–(24) imply

$$\varepsilon^k_{O(a)}(\omega) = \hat{\phi}_a + \pi_a(\omega),$$

thus defining the departure time of commodity k from the terminal $O(a)$ as the time instant at which service leg a starts to be executed by the associated service. On the other hand, if $x_a^k(\omega) = 0$, then constraints (23)–(24) are always satisfied since, due to conditions (2), $\left|\varepsilon^k_{O(a)}(\omega) - (\hat{\phi}_a + \pi_a(\omega))\right|$ is always less than or equal to $2T$. Similarly, constraints (25)–(26) ensure that the time instant $\eta^k_{D(a)}$ at which commodity k ends executing arc a equals the time instant at which the service associated with arc a (and on which k is loaded) ends executing it; otherwise, i.e., if the commodity is not transported through arc a, the constraints impose that $\left|\eta^k_{D(a)}(\omega) - \psi_a(\omega)\right|$ be lower than or equal to $2T$, which is always satisfied for the same reasons outlined for constraints (23)–(24). Finally, constraints (27) compute the delay of commodity k at its destination compared to the usual due date, and constraints (28) impose that such a delay be lower than or equal to the selected upper limit δ^k.

The travel time probability distributions can be approximated with a finite set of scenarios $\mathcal{S}$, each of which has dimension $|\mathcal{A}^{EXT}|$. Each scenario $s \in \mathcal{S}$ contains one realization of each travel time distribution and it has a probability p_s of occurrence. Given the set of scenarios $\mathcal{S}$, the expectation in the objective function (7) can be expressed as a linear function of the decision variables and the stochastic program can be formulated as a deterministic Mixed-Integer Linear Program. The scenario-based mathematical formulation is provided in the Appendix A.

5 Computational Results

This section presents the results of computational experiments conducted on a series of small to medium-sized instances utilizing commercial optimization software. In Section 5.1, details are provided on the characteristics of the instances used in the computational experiments. Moreover, Section 5.1.1 discusses stability results related to the scenario generation procedure, aiming at determining an appropriate size of the scenario set for the stochastic formulation.

The objective we have established for the computational experiments is threefold:

- To assess the efficiency of the proposed stochastic formulation in comparison to the deterministic formulation when using a commercial optimization solver. Efficiency has been measured in terms of computational time and percentage optimality gap resulting from the resolution process of the commercial solver in solving the stochastic or the deterministic instances, within a time limit of 2 hours. The numerical results related to this analysis are presented in Section 5.2;

- To evaluate the benefits of employing the stochastic formulation over the deterministic one by comparing the operational costs derived from solutions obtained using the stochastic model versus those obtained from a traditional deterministic approach. These results are discussed in Sections 5.3;

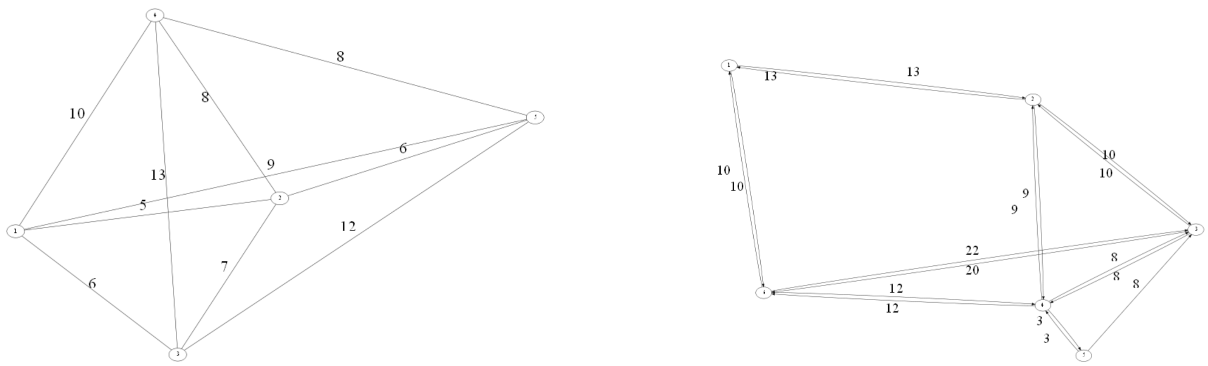
- To analyze stochastic and deterministic solutions to identify strategies that can help mitigate delays within the service network. Sections 5.4 is dedicated to these results.

Finally, in Section 5.5 we also provide an analysis focused on examining how the network design configuration obtained from the stochastic model differs from that obtained from the deterministic model, by selecting two instances as a case study. Delay probability distributions estimated for those instances are also discussed.

All the experiments were run on a computer equipped with Intel Xeon Gold 5120 CPU processors operating at 2.20 GHz and running the Ubuntu distribution of the Linux operating system. The formulation is implemented in AMPL. All the optimization models were solved with Gurobi (version 10.0.2 of the Gurobi Optimizer). Gurobi was configured to use the Network simplex algorithm, with a time limit of 2 hours.

5.1 Instances

We generated a set of test instances based on two different physical networks, inspired by the dataset employed in Boland et al. (2017); Hewitt et al. (2019). This dataset pertains to a segment of the network operated by a consolidation-based freight carrier spanning four states in the US. It encompasses various parameters such as service activation and freight routing costs, vehicle capacities, travel times, as well as service and commodity-related detailed data. Based on this dataset, the two different physical networks have been obtained by reducing the number of terminals compared to the original dataset. Precisely, we consider a first network, say $\mathcal{G}_1$, which is composed of 5 terminals (i.e., $|\mathcal{N}_1| = 5$) and 25 physical links (i.e., $|\mathcal{A}_1^{PH}| = 25$), and a second network, say $\mathcal{G}_2$, which



(a) Representation of network $\mathcal{G}_1$. (b) Representation of network $\mathcal{G}_2$.

Figure 5: Representation of the physical networks used in the experimentation.

is composed of 6 terminals (i.e., $|\mathcal{N}_2| = 6$) and 17 physical links (i.e., $|\mathcal{A}_2^{PH}| = 17$). The topologies of both physical networks are depicted in Figure 5. The usual travel time for each physical link is associated with the respective link in Figure 5a and Figure 5b. Furthermore, we considered a scheduling length of $T = 72$.

The original dataset includes only direct services. Therefore, in generating our instances we decided to add also services with one intermediate stop, for about the 15% of the total number of potential services. Regarding the activation costs of the direct services and the service capacity, which is the same for all the services, they are equal to the ones used in Boland et al. (2017) and Hewitt et al. (2019). On the other hand, the activation cost for a service with one intermediate stop is 30% less than the combined cost of the two direct services composing the route. The usual travel time of each service leg is equal to the travel time required to move along the corresponding physical link. In particular, the usual travel times range from 5 to 22 units of time. Finally, for those services having an intermediate stop, we set an idle time of 2 units of time between the end of the first leg and the beginning of the second leg.

The demand for transportation is expressed by two sets of commodities, each associated with a specific physical network. Specifically, in case of $\mathcal{G}_1$, the demand comprises 20 commodities (i.e., $|\mathcal{K}_1| = 20$), while for $\mathcal{G}_2$, it comprises 15 commodities (i.e., $|\mathcal{K}_2| = 15$). The volume of each commodity is randomly generated using a uniform distribution between 20% and 80% of the service's capacity. The availability dates are also randomly generated over the scheduling period. We consider two different delivery-time windows for both networks: a tight delivery-time window (t), with due dates about 36 units of time after the availability dates, and a loose delivery-time window (l), with due dates about 50 units of time after the availability dates. Thus, two *classes of demands*, say D-t and D-l, are defined for each physical network. Regarding the demand routing costs, they are the same used in Boland et al. (2017) and Hewitt et al. (2019), while the unit holding costs per demand volume are uniformly set to 1.

In the business-as-usual context, travel time exhibits specific characteristics, including a minimum travel time (the free running time of a service under perfect conditions, i.e., no delays), a maximum travel time (worst-case travel times excluding highly hazardous or catastrophic disturbances), and early arrivals occurring less frequently than delays. As discussed in Layeb et al. (2018), a suitable distribution for representing this type of stochastic phenomenon is the *Truncated Gamma* (TG) class of probability distributions (Chapman, 1956; Coffey and Muller, 2000). By calibrating key parameters such as mean, mode, variance, and range (defined as the difference between the maximum and minimum travel times possible), the shape of the distribution can be designed to increase rapidly to the mode (i.e., the usual travel time value), and then gradually decrease toward the maximum possible travel time, exhibiting right-skewness in its tail. This effectively captures the observed phenomena of delays occurring much more frequently than early arrivals, with delay lengths generally "not too far" from the usual travel time.

The scenario generation process involves sampling random values from TG distributions by fixing the mode and the range, and by considering two increasing levels of variability, i.e., level 1 and level 2. Level 1 denotes low variability, with a standard deviation of 0.35, while level 2 represents high variability, with a standard deviation of 0.7. The corresponding two classes of scenarios will be denoted as S-1 and S-2, depending on the considered level of variability. The range of each TG distribution is described through a lower and upper bound: the lower bound is equal to the mode minus 5% of the mode, while the upper bound is equal to the mode plus 35% of the mode. Figure 6 shows an example of the two TG distributions associated with a service leg with a mode equal to 8. We assumed independent sampling, implying independent travel time random variables. Note that this assumption only impacts on the scenario generation procedure, not on the performance of the scenario-based model.

Lastly, we considered two levels of allowed delays for both service departure times and commodity delivery times, specifically 2 and 6 units of time (AD-1 and AD-2, respectively). We also considered two levels of costs for service delays, commodity late arrivals at destinations, and outsourcing, denoted by P-1 and P-2, where the values in P-2 double those in P-1. The above-mentioned attributes are recapped in Table 2.

Table 2: Attributes of the classes of instance.

Physical Network	$\mathcal{G}_1$, $\mathcal{G}_2$
Delivery-time windows	D-t, D-l
Variability levels of the TG	S-1, S-2
Allowed delays for service and commodities	AD-1, AD-2
Penalties for delays and outsourcing	P-1, P-2

In summary, for each network topology, i.e., $\mathcal{G}_1$ and $\mathcal{G}_2$, we have generated 16 *classes of instances* based on the combination of the two demand classes D-t and D-l, the two levels of variability of the TG S-1 and S-2, the two levels of allowed delays for service departure times and commodity delivery time (AD-1 and AD-2), and the two levels of costs (P-1 and P-2). For each of these 16 classes of instances, we have generated 10 instances randomly. Thus, a total of 320 instances have been generated and solved.

5.1.1 In-sample and out-of-sample stability

We conducted in-sample and out-of-sample stability verification to evaluate the accuracy of our scenario generation procedure and the representativeness of the generated scenarios, and, thus, avoid introducing bias in the results of the optimization model. In fact, once both in-sample and out-of-sample stabilities are verified, we can confidently ensure that the solutions obtained by solving the stochastic formulation are unaffected by any bias introduced by the specific scenario set used in the resolution process (see Kaut et al., 2007; King and Wallace, 2012, for additional details).

We conducted stability tests for both network topologies, i.e., $\mathcal{G}_1$ and $\mathcal{G}_2$, by considering instances belonging to S-2 (the highest variability level), D-t and D-l (tight and loose

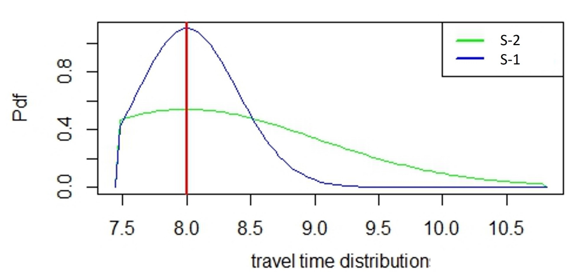


Figure 6: Travel time distribution of TG with mode 8 for both variability levels.

Table 3: In-sample stability results for the scenario-generation procedure.

Network $\mathcal{G}_1$						
	Avg. Relative Range		Avg. Relative Mean Deviation		Avg. Relative Median Deviation	
	AD-1	AD-2	AD-1	AD-2	AD-1	AD-2
D-t	4.88%	3.90%	1.57%	1.21%	0.81%	0.71%
D-l	5.99%	3.30%	1.94%	1.02%	1.33%	0.50%

Network $\mathcal{G}_2$						
	Avg. Relative Range		Avg. Relative Mean Deviation		Avg. Relative Median Deviation	
	AD-1	AD-2	AD-1	AD-2	AD-1	AD-2
D-t	9.64%	5.43%	3.60%	1.92%	3.46%	1.77%
D-l	6.43%	3.84%	2.27%	1.31%	2.41%	1.32%

delivery-times), AD-1 and AD-2 (2 and 6 units of time of delay for service departure times and commodity delivery time), and P-2 (the highest level of penalty). Approximating a probability distribution using a set of scenarios introduces noise, which decreases as the number of scenarios increases. However, this also raises the complexity of obtaining solutions. Thus, there is a trade-off between stability and problem size. Preliminary tests indicated that scenario sets of size 20 yielded satisfactory results, meeting our goals for solution accuracy and acceptable computational time.

Regarding in-sample stability, each instance from the aforementioned classes was solved 10 times, each with a different scenario set of size 20. The best objective function values, obtained within a 2-hour time limit, were collected and analyzed. For each instance, three statistical indicators were used for the analysis: the relative range, i.e., the difference between the highest and the lowest objective function values, the relative mean deviation, and the relative median deviation (note that, although the mean is a more classical indicator, it may not represent the value of a possible solution as the median does). Table 3 reports average aggregated results related to three mentioned statistical indicators, namely the average range (column 1), the average relative mean deviation (column 2), and the average relative median deviation (column 3) across instances of the same classes, for $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively.

Table 4: Out-of-sample stability results for the scenario-generation procedure.

Network $\mathcal{G}_1$						
	Avg. Relative Range		Avg. Relative Mean Deviation		Avg. Relative Median Deviation	
	AD-1	AD-2	AD-1	AD-2	AD-1	AD-2
D-t	3.59%	1.85%	1.38%	0.74%	1.32%	0.85%
D-l	4.57%	1.03%	1.71%	0.34%	1.30%	1.12%
Network $\mathcal{G}_2$						
	Avg. Relative Range		Avg. Relative Mean Deviation		Avg. Relative Median Deviation	
	AD-1	AD-2	AD-1	AD-2	AD-1	AD-2
D-t	5.43%	1.98%	2.41%	0.74%	4.04%	0.60%
D-l	4.04%	0.69%	1.65%	0.24%	2.15%	0.35%

Regarding out-of-sample stability, the solutions obtained during the in-sample testing were evaluated by approximating the "true" travel time probability distributions resorting to a 100-scenario-sized set, generated from the same TG distribution used to construct the scenario sets for the optimization process. Table 4 shows average aggregated results related to the three above-mentioned statistical indicators across instances of the same classes, for $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively.

It is worth noting that the reported results pertain to the highest variability level, representing the most volatile situations. In less variable cases, the obtained values are even lower. Consequently, all the results discussed in the next sessions are obtained by solving stochastic formulations with a scenario size of 20.

5.2 Efficiency of the stochastic approach

In this section, we evaluate the performance of the proposed stochastic formulation, referred to as STT, in terms of average computational time (in seconds) and average percentage gap (with a time limit of two hours), compared to the deterministic counterpart of the formulation, called DET. Specifically, DET is characterized by a single scenario defined in terms of the usual travel times. We analyze the performance across all the classes of instances and for the two network topologies, $\mathcal{G}_1$ and $\mathcal{G}_2$. The results of the analysis are reported in Table 5. These results as well as all the following results are averages over the 10 instances belonging to the same class of instances.

Gurobi solved all the instances in a very short time when considering DET, requiring only a few seconds on average to obtain optimal solutions regardless of the network topology considered. However, it appears that looser delivery time windows render the problem slightly more challenging to solve, likely due to the increased number of feasible itineraries for the commodities. Conversely, and as expected, a significantly longer computational time is required to find solutions in the case of STT. Specifically, the parameter combination S-1 (low variability), D-t (tight delivery time window), AD-1 (2 units of time allowed delay), and P1 (low penalty) defines instances that can still be solved to optimality by Gurobi in slightly more than one hour on average, no matter what network topology is used. On the other hand, the other parameter combinations render the problem marginally harder to solve: Gurobi can compute near-optimal solutions, with generally low optimality gaps, by reaching in some cases the time limit of 2 hours imposed. The most challenging parameter combination for both topologies appears to be S-2 (high variability), D-l (loose delivery time window), AD-2 (6 units of time allowed delay), and P2 (high penalty). In the case of the network topology $\mathcal{G}_1$ this combination determines the highest optimality gaps, on average, as well as the highest computational time, i.e., 2 hours.

5.3 Evaluation Analysis: Benefits of Stochastic Formulation

This analysis aims to quantify the benefits of explicitly incorporating time-stochasticity into the formulation, rather than relying on a traditional deterministic approach. Such benefits are discussed in terms of two types of costs: *i*) the network setup costs, i.e., the costs incurred by selecting the services; and *ii*) the estimated full costs, providing information on the performance of the selected service network in terms of network setup costs, routing costs, costs for delays in service operations and commodities deliveries, and outsourcing.

Given a solution obtained through either DET or STT with a scenario set of size 20, the full costs are determined using a simulation-like procedure. This procedure involves solving the second stage of the mathematical model by fixing the services selected in DET or STT and using a scenario set of size 100, generated from the same scenario generation procedure used to construct the scenario sets for the optimization process of STT. This analysis encompasses all classes of instances and both network topologies.

Table 5: Performance of STT vs DET.

Network $\mathcal{G}_1$

			P-1		P-2	
			Time (sec.)	Opt. Gap	Time (sec.)	Opt. Gap
D-t	AD-1	Det	1.93	0%	1.86	0%
		S-1	4301	0%	5293	0.%
		S-2	5278	0.17%	4952	0.06%
D-t	AD-2	Det	3.62	0%	3.46	0%
		S-1	6751	0.29%	7031	0.48%
		S-2	7014	0.68%	7200	1.75%
D-l	AD-1	Det	9.70	0%	8.72	0%
		S-1	7200	0.93%	7200	0.93%
		S-2	6743	0.36%	6678	0.83%
D-l	AD-2	Det	13.20	0%	10.49	0%
		S-1	7200	1.91%	7200	1.94%
		S-2	7200	1.59%	7200	2.54%

Network $\mathcal{G}_2$

			P-1		P-2	
			Time (sec.)	Opt. Gap	Time (sec.)	Opt. Gap
D-t	AD-1	Det	1.20	0%	1.60	0%
		S-1	1287	0%	1404	0%
		S-2	2523	0%	3400	0.11%
D-t	AD-2	Det	2.01	0%	2.46	0%
		S-1	3313	0.28%	4063	0.94%
		S-2	4429	0.55%	5719	2.09%
D-l	AD-1	Det	1.83	0%	1.79	0%
		S-1	2222	0%	2641	0%
		S-2	3502	0.01%	4348	0.26%
D-l	AD-2	Det	2.58	0%	2.36	0%
		S-1	5757	0.25%	6343	0.46%
		S-2	6854	0.54%	6731	1.30%

The obtained results are presented for both network topologies in Table 6. The table shows the percentage increase or decrease of the objective function values (Obj. Value), setup costs, and full costs of the stochastic solutions with respect to the deterministic ones.

The objective function values obtained through the stochastic approach are higher than those achieved using the deterministic approach. This outcome is expected because the deterministic approach assumes volatility-free conditions and thus no delays exceeding planned movements.

Of particular interest are the setup costs of DET and STT for both network topologies. Specifically, for network topology $\mathcal{G}_1$, STT consistently yields lower setup costs compared to its deterministic counterpart DET. On the other hand, for network topology $\mathcal{G}_2$ this is not always true. In fact, the stochastic solutions may incur higher setup costs, under certain parameter configurations, than their deterministic counterpart. That is to say, more services are activated in the solutions determined by STT than in ones computed by DET to meet the demand within the agreed service quality. This especially occurs when solving STT under the penalty cost P-1, showing a maximum increase in setup costs for the most volatile case and tight delivery time windows (i.e., for the configuration P-1, D-t, S-2).

Nevertheless, across all classes of instances and topology networks, STT always exhibits lower full costs than DET (recall full costs include service activation costs, commodity routing costs, costs for delays and outsourcing). This underscores the advantage of explicitly incorporating the stochastic nature of travel times in tactical planning models, which can mitigate or reduce the economic impacts and consequences of uncertainty, despite possible initial higher setup costs in a few cases. For both network topologies, the greatest full cost reduction of STT with respect to DET is observed when the variability of travel time distributions is high and the reliability is maximally enforced in the formulation through the penalty cost P-2. Precisely, the highest reduction in full costs is observable for the network topology $\mathcal{G}_2$ when considering D-l and AD-1.

5.4 Structural Analysis: Reducing Delay Risk

The purpose of this analysis is to identify the features that stochastic solutions exploit to hedge against time uncertainty, by considering all classes of instances and both network topologies.

Table 6: Cost comparison of STT vs DET.

Network $\mathcal{G}_1$								
			P-1			P-2		
			Obj. Value	Setup Cost	Full Cost	Obj. Value	Setup Cost	Full Cost
		Det	-	-	-	-	-	
D-t	AD-1	S-1	+5.99%	-11.14%	-2.48%	+11.27%	-10.64%	-5.38%
		S-2	+18.26%	-11.57%	-11.08%	+32.36%	-10.67%	-18.27%
		Det	-	-	-	-	-	-
D-t	AD-2	S-1	+6.17%	-10.34%	-2.56%	+11.46%	-10.47%	-4.62%
		S-2	+17.95%	-11.80%	-11.11%	+32.33%	-11.99%	-16.32%
		Det	-	-	-	-	-	-
D-l	AD-1	S-1	+4.60%	-3.75%	-4.51%	+8.70%	-9.73%	-6.48%
		S-2	+14.86%	-7.59%	-4.75%	+26.38%	-14.09%	-17.87%
		Det	-	-	-	-	-	-
D-l	AD-2	S-1	+4.62%	-1.41%	-4.14%	+8.70%	-10.70%	-9.00%
		S-2	+13.97%	-7.94%	-9.11%	+25.96%	-16.33%	-21.99%

Network $\mathcal{G}_2$								
			P-1			P-2		
			Obj. Value	Setup Cost	Full Cost	Obj. Value	Setup Cost	Full Cost
		Det	-	-	-	-	-	
D-t	AD-1	S-1	+9.17%	+7.28%	-9.00%	+17.81%	+0.95%	-10.22%
		S-2	+37.30%	+14.76%	-21.77%	+71.63%	+6.71%	-23.99%
		Det	-	-	-	-	-	-
D-t	AD-2	S-1	+6.00%	+3.83%	-6.91%	+13.21%	-3.00%	-13.76%
		S-2	+14.07%	+4.59%	-18.98%	+28.61%	-1.90%	-30.55%
		Det	-	-	-	-	-	-
D-l	AD-1	S-1	+4.54%	+1.96%	-18.41%	+8.77%	-1.01%	-23.79%
		S-2	+11.96%	+2.51%	-27.05%	+20.91%	-0.42%	-35.08%
		Det	-	-	-	-	-	-
D-l	AD-2	S-1	+4.52%	-0.16%	-16.43%	+9.02%	-1.88%	-24.03%
		S-2	+10.59%	-2.91%	-24.57%	+19.86%	-4.55%	-33.81%

Table 7 illustrates the composition of the service network of the solutions obtained through DET and STT. In particular, for all parameter combinations, the table presents the percentage difference in the total number of services activated in DET and in STT, as well as disaggregated information in terms of number of direct and not direct activated services.

In all problem instances and network topology $\mathcal{G}_1$, STT operates fewer services (considering the sum of direct and not direct) compared to DET, sharing however some activated services. This trend is also observable in the majority of the classes of instances related to network topology $\mathcal{G}_2$, with exceptions for D-t, such as when AD-1 or AD-2 and P-1 are used, where the number of activated services in STT is higher than in DET. The reduction in the total number of activated services may be primarily due to the fact that each active service entails the risk of delays and thus associated costs. Regarding the number of active direct services, in $\mathcal{G}_2$ it is always higher in STT than in DET. A notable remark is the consistent reduction in the activation of non-direct services in the STT solutions, no matter the network topology considered, which are particularly vulnerable to delays. When delays occur on the first leg of these services, they often lead to late arrivals at the second stop, unless the travel time for the second leg is significantly shorter than the usual one, thereby mitigating the delay. Given the assumed probability distributions, complete absorption of delays is unlikely, resulting in higher risks and costs associated with operating one-stop services. Consequently, STT tends to prioritize more expensive direct connections to less costly multi-stop routes, to try to reduce the potential for additional costs during operations. This behaviour may also account for the increase in the number of activated direct services observed in the classes of instances related to $\mathcal{G}_2$, as substituting a non-direct service requires activating two direct services. Note that, in the latter cases, despite the increased number of active services, the operational costs associated with the STT solutions are lower compared to DET ones, as discussed in the previous section. Generally, STT tends to design service networks composed only of direct services which are essential to meet the demand, by replacing non-direct services with direct services as necessary.

Table 8 illustrates some features relative to the routing of the commodities in the DET and the STT solutions, given the selected service network. In particular, for all parameter combinations, the table presents the percentage difference in the total number of direct transportation for commodities (i.e., commodities moved directly from their origins to their destinations without passing through intermediate terminals), the percentage difference in early arrivals of commodities at their destination, and the percentage

Table 7: Service network features: STT vs DET.

Network $\mathcal{G}_1$

			P-1			P-2		
			Total Services	Direct Services	Not direct Services	Total Services	Direct Services	Not direct Services
		Det	-	-	-	-	-	
D-t	AD-1	S-1	-11.09%	-11.31%	-10.75%	-11.50%	-7.69%	-22.87%
		S-2	-13.63%	-4.54%	-37.15%	-14.63%	+3.83%	-52.85%
		Det	-	-	-	-	-	-
D-t	AD-2	S-1	-10.28%	-10.48%	-9.08%	-11.09%	-8.55%	-20.58%
		S-2	-13.49%	-5.81%	-32.27%	-15.35%	-0.13%	-50.24%
		Det	-	-	-	-	-	-
D-l	AD-1	S-1	-4.67%	-0.49%	-12.83%	-10.81%	-6.02%	-20.83%
		S-2	-9.92%	+0.38%	-34.00%	-17.17%	-3.60%	-49.67%
		Det	-	-	-	-	-	-
D-l	AD-2	S-1	-1.92%	+0.38%	-6.43%	-12.01%	-5.66%	-22.00%
		S-2	-9.45%	-2.73%	-27.69%	-19.01%	-6.39%	-45.64%

Network $\mathcal{G}_2$

			P-1			P-2		
			Total Services	Direct Services	Not direct Services	Total Services	Direct Services	Not direct Services
		Det	-	-	-	-	-	
D-t	AD-1	S-1	+6.34%	+10.26%	-5.00%	-0.57%	+5.85%	-19.33%
		S-2	+11.22%	+26.01%	-34.02%	+2.53%	+19.83%	-55.69%
		Det	-	-	-	-	-	-
D-t	AD-2	S-1	+2.65%	+7.69%	-11.17%	-4.67%	+2.60%	-21.08%
		S-2	+1.67%	+14.10%	-33.02%	-5.08%	+8.33%	-47.75%
		Det	-	-	-	-	-	-
D-l	AD-1	S-1	+0.30%	+7.82%	-17.11%	-2.87%	+5.22%	-29.71%
		S-2	-1.02%	+14.55%	-43.54%	-5.06%	+14.99%	-61.60%
		Det	-	-	-	-	-	-
D-l	AD-2	S-1	-1.94%	+6.11%	-22.25%	-4.09%	+5.35%	-35.74%
		S-2	-5.65%	+6.44%	-37.77%	-8.47%	+8.55%	-56.55%

difference in routing costs. In STT, commodity itineraries slightly deviate from the typical consolidation-based itineraries, where different commodities share the capacity of a single service for most of their journeys, often passing through several intermediate stops where they wait idly before reaching their destination. In contrast, the STT solutions show an increase in the percentage of direct and dedicated itineraries for commodities compared to the deterministic solutions. There is also a noticeable decrease in just-in-time arrivals of commodities at their destinations with respect to what happens in the deterministic solutions, with more commodities arriving well in advance with respect to their due dates. In fact, the percentage of early arrivals is greater across all cases. Despite different service networks having been designed, routing costs are nearly the same in all the STT and DET solutions.

Ultimately, Table 9 illustrates the benefits in terms of reliability in service operations and punctuality of freight delivery at destinations in the solutions obtained through STT. Specifically, for all parameter combinations, the table presents the percentage reduction in delays in operating the service network, as well as the percentage reduction in delivery delays for the STT solutions compared to the DET ones. Regarding the service network, delays are reduced in STT compared to DET by at least 12% and up to 45% for the majority of the classes of instances across both network topologies. However, in topology $\mathcal{G}_2$ and for two problem classes, namely D-t, AD-2, and P-1, the delays in service operations are slightly higher in STT compared to DET. In terms of deliveries, delays are at least halved in the STT solutions compared to the DET ones. Furthermore, in some cases, STT almost eliminates delivery delays, achieving reductions of over 90%. This is evident in the classes of instances such as D-t, AD-1 or AD-2, P-1 and S-2 for $\mathcal{G}_1$ or D-t, AD-2, P-2 and S-1 or S-2 for $\mathcal{G}_2$.

Lastly, it is important to note that, across all classes of instances and for both network topologies, increasing the penalty from P-1 to P-2, thereby heightening in the service network the need to hedge against fluctuations in travel time, enhances the reliability of service operations. However, this increase in reliability comes at the expense of delivery punctuality, which slightly decreases when using P-2 instead of P-1.

The features above clearly demonstrate that the stochastic formulation of the problem is essential for achieving the observed benefits of delay reductions. Such improvements in service reliability and delivery punctuality would not have been determined with a traditional time-deterministic formulation, underscoring the value of the stochastic approach when quality targets are of interest. Furthermore, such an enhancement in service

Table 8: Commodity routing features: STT vs DET.

Network $\mathcal{G}_1$

			P-1			P-2		
			Direct Itiner.	Early Arrivals	Routing Costs	Direct Itiner.	Early Arrivals	Routing Costs
		Det	-	-	-	-	-	
D-t	AD-1	S-1	+5.68%	+8.24%	+0.07%	+6.14%	+16.23%	-0.48%
		S-2	+17.65%	+ 15.97%	+1.95%	+23.33%	+28.71%	+3.13%
		Det	-	-	-	-	-	-
D-t	AD-2	S-1	+5.65%	+5.39%	-0.09%	+5.85%	+7.04%	-0.10%
		S-2	+15.88%	+14.59%	+0.60%	+22.74%	+17.36%	+3.30%
		Det	-	-	-	-	-	-
D-l	AD-1	S-1	+6.64%	+19.08%	-1.06%	+3.77%	+5.49%	+0.74%
		S-2	+19.27%	+10.99%	+1.34%	+17.35%	+13.83%	+4.63%
		Det	-	-	-	-	-	-
D-l	AD-2	S-1	+3.76%	+11.20%	-1.29%	+2.82%	+8.54%	-0.15%
		S-2	+16.53%	+15.71%	-0.45%	+19.18%	+28.63%	+2.13%

Network $\mathcal{G}_2$

			P-1			P-2		
			Direct Itiner.	Early Arrivals	Routing Costs	Direct Itiner.	Early Arrivals	Routing Costs
		Det	-	-	-	-	-	
D-t	AD-1	S-1	-0.41%	+21.90%	-0.03%	+1.02%	+14.32%	0%
		S-2	+7.89%	+26.26%	+2.27%	+9.50%	+26.67%	+2.25%
		Det	-	-	-	-	-	-
D-t	AD-2	S-1	-1.39%	+7.05%	+0.54%	-1.08%	+14.14%	+0.33%
		S-2	+6.93%	+14.96%	+0.97%	+6.52%	+19.06%	+0.95%
		Det	-	-	-	-	-	-
D-l	AD-1	S-1	+4.04%	+18.79%	+1.48%	+4.26%	+14.57%	+0.79%
		S-2	+7.09%	+17.16%	+4.14%	+3.95%	+23.55%	+4.85%
		Det	-	-	-	-	-	-
D-l	AD-2	S-1	+3.98%	+2.27%	+1.24%	+5.82%	+6.95%	+1.70%
		S-2	+6.56%	+11.39%	+2.09%	+4.71%	+16.22%	+3.70%

Table 9: Delay features: STT vs DET.

Network $\mathcal{G}_1$						
			P-1		P-2	
			Service Network Delays	Total Commodity Delays	Service Network Delays	Total Commodity Delays
		Det	-	-	-	-
D-t	AD-1	S-1	-16.48%	-59.57%	-25.28%	-59.56%
		S-2	-43.06%	-96.27%	-45.53%	-83.07%
		Det	-	-	-	-
D-t	AD-2	S-1	-21.87%	-64.66%	-26.51%	-58.52%
		S-2	-37.97%	-95.49%	-42.47%	-45.93%
		Det	-	-	-	-
D-l	AD-1	S-1	-18.83%	-49.50%	-21.89%	-39.26%
		S-2	-34.25%	-79.75%	-39.94%	-95.99%
		Det	-	-	-	-
D-l	AD-2	S-1	-14.67%	-66.39%	-27.57%	-66.41%
		S-2	-25.97%	-79.99%	-38.36%	-69.73%

Network $\mathcal{G}_2$						
			P-1		P-2	
			Service Network Delays	Total Commodity Delays	Service Network Delays	Total Commodity Delays
		Det	-	-	-	-
D-t	AD-1	S-1	-12.57%	-72.99%	-25.47%	-51.79%
		S-2	-10.72%	-69.06%	-24.30%	-69.48%
		Det	-	-	-	-
D-t	AD-2	S-1	+7.16%	-85.19%	+0.77%	-93.90%
		S-2	-10.32%	-90.90%	-17.76%	-94.12%
		Det	-	-	-	-
D-l	AD-1	S-1	-20.74%	-79.99%	-25.75%	-80.00%
		S-2	-32.89%	-79.83%	-41.23%	-78.20%
		Det	-	-	-	-
D-l	AD-2	S-1	-22.59%	-87.78%	-30.80%	-79.99%
		S-2	-27.68%	-88.48%	-36.96%	-89.52%

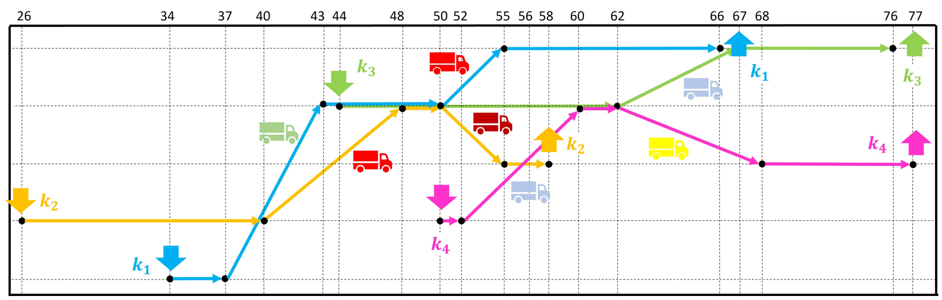
network quality comes with a reduction in overall operating costs, as shown by the results in Table 6.

5.5 Delay Distributions and Routes

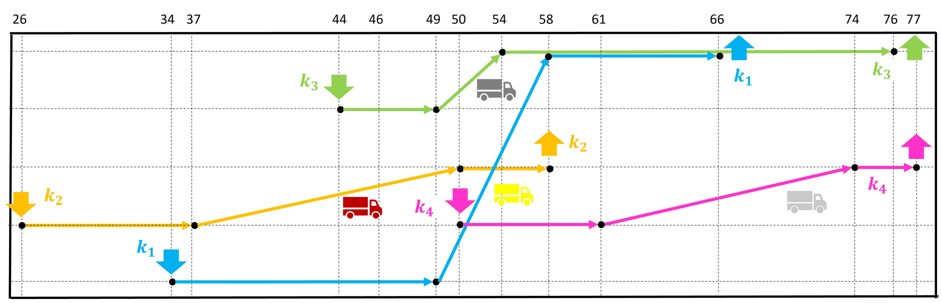
The goal of this final analysis is to examine how the network design configuration of STT differs from that of DET, by giving a closer look at some of the instances solved. Precisely, we analyze the structure and the performance of one instance for each of the two network topologies, namely one instance with parameter combination AD-2, D-t, S-2, and P-2 for $\mathcal{G}_1$ and one instance with parameter combination AD-2, D-l, S-2, and P-1 for $\mathcal{G}_2$.

Firstly, we present a comparison between the DET and the STT itineraries, by focusing on a subset of the commodities rather than showing the complete network, for clarity and readability motivations. The comparison of the network structure is illustrated using the time chart methodology employed in Section 4.1 for Example 1. In this chart, the horizontal axis represents time and the vertical axis represents geographical location, i.e., terminals. The availability and due dates of commodities are indicated by solid arrows, itineraries of commodities are shown using coloured arrows, and activated services are represented by the symbol of a truck. Additionally, we provide statistics related to the distribution of delays in service operations.

Regarding $\mathcal{G}_1$, DET displays, in general, characteristics typical of consolidation-based transportation networks, where different commodities share the capacity of single services for most of their journeys, passing through intermediate stops, before arriving at their destination. Furthermore, one-stop services are usually favoured when possible rather than (more costly) direct services, to lower fixed costs. Figure 7a shows the itineraries of four commodities and the associated services to move them in DET. In particular, five services are activated, two of them (indicated through a red and blue truck) with one intermediate stop. Three out of four commodities required transhipment to arrive at the destination (e.g., k_2 is moved by the green and the red service being transhipped at terminal 2). Figure 7b shows the itineraries of the same four commodities and the associated services to move them in STT, in which all commodities are moved directly from their origins to their destinations through 4 direct and dedicated services. The use of direct and dedicated services likely helps to reduce delays that arise from the need to synchronize multiple shipments, which require consolidation at specific terminals before they can continue their journeys together.

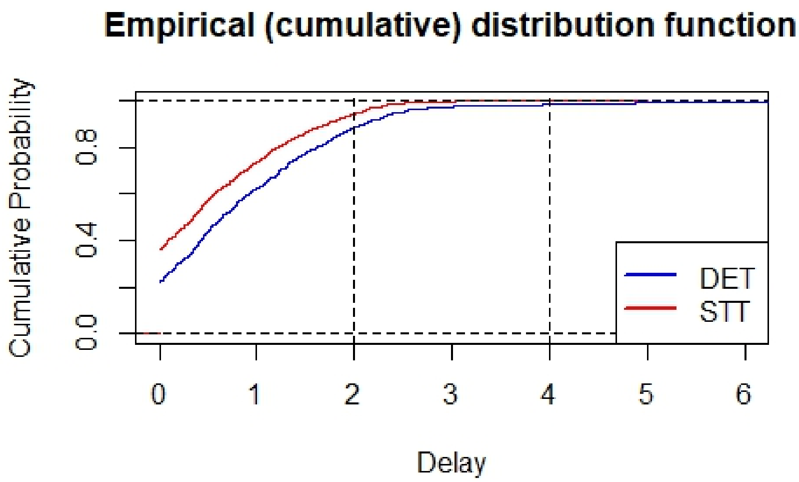


(a) Deterministic setting in $\mathcal{G}_1$



(b) Stochastic setting in $\mathcal{G}_1$

Figure 7: Comparison of the itineraries for five commodities in $\mathcal{G}_1$, in deterministic and stochastic settings.



(a) Empirical cumulative distribution of the delay for an instance in $\mathcal{G}_1$.

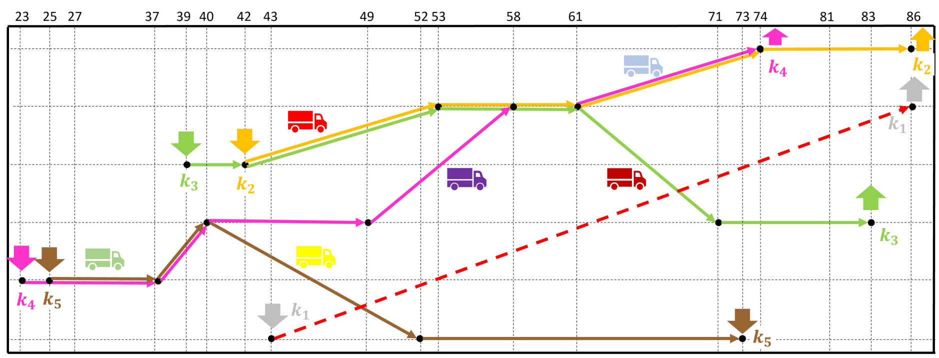
PDF	Min	Max	Mean	St.Dev.	Median	Kurtosis	Skewness
DET	0	640.32	27.73	38.49	19.03	113.70	7.95
STT	0	129.32	18.03	21.10	10.40	0.62	1.13

(b) Descriptive statistics of the probability distribution of the delay for an instance in $\mathcal{G}_1$.

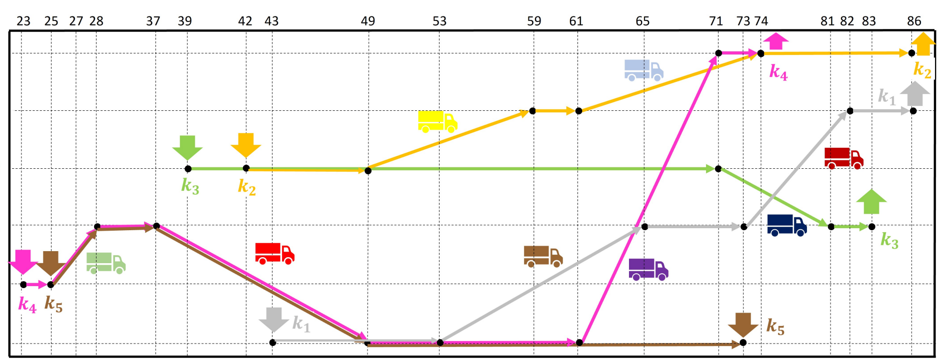
Figure 8: Characteristics of the probability distribution of delays.

Figure 8 reports some statistics related to the delays in service operations of the above-considered instance. Specifically, Figure 8a reports the empirical cumulative (probability) distribution function (CDF) of the delays of the service network for DET (blue function) and STT (red function). The CDF of STT exhibits a higher steepness towards its maximum value compared to the CDF of DET. Additionally, the CDF of STT consistently lies above the CDF of DET. This highlights a trend towards shorter delays for STT, which are thus more likely to be observed. In contrast, the broader support and extended tail of the CDF for DET indicate a significant proportion of delays extending beyond the range observed in the red CDF. Figure 8b reports some statistical indicators of the probability density function (PDF) of DET and STT (we refer to Moore and McCabe, 1989; Newbold et al., 2013, for a detailed explanantion of the mentioned indicators).

Figure 9a and Figure 9b show the itineraries of five commodities and the associated services to move them in DET and STT, respectively, referring to network topology $\mathcal{G}_2$.



(a) Deterministic setting in $\mathcal{G}_2$



(b) Stochastic setting in $\mathcal{G}_2$

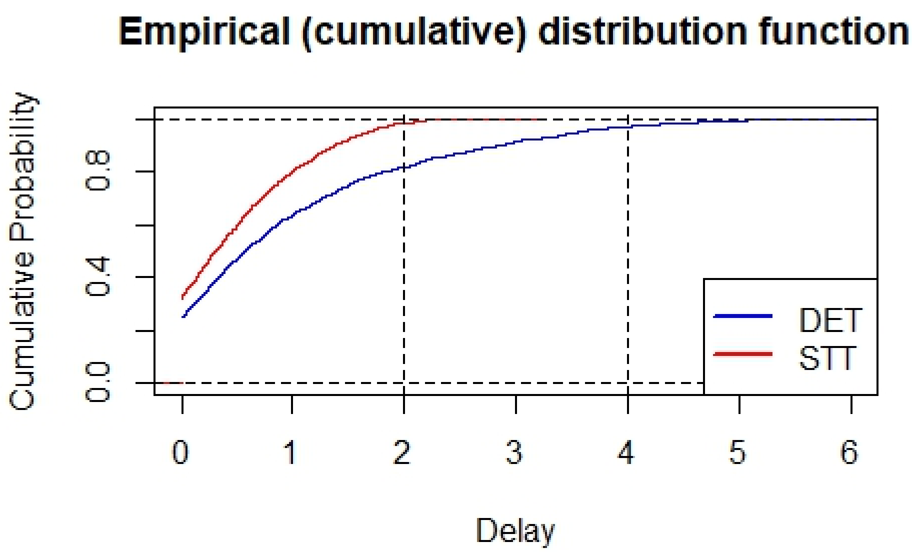
Figure 9: Comparison of the itineraries for five commodities in $\mathcal{G}_2$ under deterministic and stochastic settings.

As mentioned in Section 5.1 and as opposed to $\mathcal{G}_1$, not all connections among terminals are available in network topology $\mathcal{G}_2$. This means that for some commodities, different potential itineraries are not available to move them from their origins to their destinations, being obliged to be moved along certain links despite likely experiencing delays. Comparing the solutions in DET and STT, this occurs, for instance, with commodity k_2, where the physical link on which it is moved is the same in both DET and STT. Nevertheless, in DET, its first movement is performed by a non-direct service (i.e., the one indicated with a red truck), which is more vulnerable to delays. This is replaced in STT by a direct service (i.e., the one indicated with a yellow truck). Note also the already highlighted behavior regarding commodity k_3, being moved by two services in DET and by one direct and dedicated service in STT. The solution in DET accounts for a just-in-time arrival of k_4 at the destination, which is avoided in STT by defining a more reliable itinerary against time fluctuations. Finally, note that commodity k_1 is outsourced in DET (indicated by a dashed red line), while it is moved by two services in STT.

Figure 10a reports the empirical cumulative distribution function (CDF) of the service network delays for DET (blue function) and STT (red function) for the instance under consideration. Comparing the two CDFs, the higher steepness towards its maximum value of the CDF of STT and the broader support and extended tail of the CDF for DET are even more evident. Finally, Figure 10b presents some statistical indicators of the probability density function (PDF) for both DET and STT.

6 Conclusions

In this paper, we addressed the Continuous-Time Service Network Design problem with Stochastic Travel Times, extending the classical SSND problem to consider business-as-usual fluctuations in travel times, aiming at reducing the negative impact of delays on the feasibility and profitability of transportation plans. In this setting, we assume service selection is made at the planning stage, before observing the actual travel time realizations, only considering probability distributions. Once these decisions are made, we assume more precise estimations of travel times based on system conditions are available to make additional decisions regarding service operations and commodity movements. Additionally, penalties for delays of services at stops with respect to the schedule and of commodities at their destination with respect to due dates are also computed.



(a) Empirical cumulative distribution of the delay for an instance in $\mathcal{G}_2$.

PDF	Min	Max	Mean	St.Dev.	Median	Kurtosis	Skewness
DET	0	183.80	30.35	36.03	17.38	1.79	1.48
STT	0	82.32	15.20	17.23	9.16	0.40	1.10

(b) Descriptive statistics of the probability distribution of the delay for an instance in $\mathcal{G}_2$.

Figure 10: Characteristics of the probability distribution of delays.

The problem has been formulated as a two-stage stochastic programming model encompassing planning (i.e., service selection) and recourse phases (i.e., postponement of service departures, commodities itineraries, and outsourcing). We emphasize that such recourse actions have never been considered together in an SSND formulation with stochastic travel times. Our problem formulation utilizes an Stochastic-Aware Service-Leg Network, which models time continuously and compactly, mitigating the size issues associated with traditional time-space networks.

Through extensive experimentation with realistic small-to-medium size instances, we assessed the complexity and benefits of our stochastic formulation compared to the deterministic one. The results demonstrated that our approach effectively mitigates delays and improves operational costs, highlighting specific features to hedge against time fluctuations appearing in stochastic solutions.

Future research avenues include extending this formulation within a deterministic setting, including, for instance, additional management issues such as empty repositioning of resources (e.g., vehicles or containers) or exploring other sources of uncertainty related to demand (e.g., volume or availability dates) or operation times at terminals. Finally, developing sophisticated and efficient solution methods will be essential for handling larger instances.

Funding

Financed by the European Union-NextGenerationEU, National Sustainable Mobility Center CN00000023, Italian Ministry of University and Research Decree n. 1033-17/06/2022, Spoke 10.

Acknowledgments

While working on the project, the second 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 Montrèal. He gratefully acknowledges the financial support provided by the Natural Sciences and Engineering Council of Canada, through its Discovery grant program, and of the Fonds de recherche du Québec through their infrastructure grants.

References

Andersen, J., Crainic, T. G., and Christiansen, M. (2009a). Service network design with asset management: Formulations and comparative analyses. *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.

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.

Birge, J. R. and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Boland, N., Hewitt, M., Marshall, L., and Savelsbergh, M. (2017). The continuous-time service network design problem. *Operations research*, 65(5):1303–1321.

Boland, N., Hewitt, M., Marshall, L., and Savelsbergh, M. (2019). The price of discretizing time: a study in service network design. *EURO Journal on Transportation and Logistics*, 8(2):195–216.

Chapman, D. G. (1956). Estimating the parameters of a truncated gamma distribution. *The Annals of Mathematical Statistics*, pages 498–506.

Coffey, C. S. and Muller, K. E. (2000). Properties of doubly-truncated gamma variables. *Communications in Statistics-Theory and Methods*, 29(4):851–857.

Crainic, T. G. (2000). Service network design in freight transportation. *European journal of operational research*, 122(2):272–288.

Crainic, T. G. (2024). Consolidation-based transportation planning: The service network design methodology. In *CIRRELT-2024-12*, pages 1–59. CIRRELT.

Crainic, T. G., Gendreau, M., and Gendron, B. (2021). *Network design with applications to transportation and logistics*. Springer.

Crainic, T. G. and Hewitt, M. (2021). Service network design. In *Network Design with Applications to Transportation and Logistics*, pages 347–382. Springer.

Crainic, T. G., Hewitt, M., Toulouse, M., and Vu, D. M. (2016). Service network design with resource constraints. *Transportation Science*, 50(4):1380–1393.

Crainic, T. G., Hewitt, M., Toulouse, M., and Vu, D. M. (2018). 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. *Handbooks in operations research and management science*, 14:467–537.

Demir, E., Burgholzer, W., Hrušovskỳ, M., Arıkan, E., Jammernegg, W., and Van Woensel, T. (2016). A green intermodal service network design problem with travel time uncertainty. *Transportation Research Part B: Methodological*, 93:789–807.

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. and Lehuédé, F. (2023). New formulations for the scheduled service network design problem. *Transportation Research Part B: Methodological*, 172:117–133.

Hewitt, M., Rei, W., and Wallace, S. W. (2021). Stochastic network design. *Network Design with Applications to Transportation and Logistics*, pages 283–315.

Hoff, A., Lium, A.-G., Løkketangen, A., and Crainic, T. G. (2010). A metaheuristic for stochastic service network design. *Journal of Heuristics*, 16:653–679.

Hrušovskỳ, M., Demir, E., Jammernegg, W., and Van Woensel, T. (2018). Hybrid simulation and optimization approach for green intermodal transportation problem with travel time uncertainty. *Flexible Services and Manufacturing Journal*, 30:486–516.

Kaut, M., Vladimirou, H., Wallace, S. W., and Zenios, S. A. (2007). Stability analysis of portfolio management with conditional value-at-risk. *Quantitative Finance*, 7(4):397–409.

King, A. J. and Wallace, S. W. (2012). *Modeling with stochastic programming*. Springer Science & Business Media.

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.

Lanza, G., Crainic, T. G., Passacantando, M., and Scutellà, M. G. (2024). A continuous time physical graph based formulation to scheduled service network design. In *International Conference on Optimization and Decision Science*, 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. In *International Conference on Computational Logistics*, pages 401–416. Springer.

Lanza, G., Crainic, T. G., Rei, W., and Ricciardi, N. (2021). Scheduled service network design with quality targets and stochastic travel times. *European Journal of Operational Research*, 288(1):30–46.

Layeb, S. B., Jaoua, A., Jbira, A., and Makhlouf, Y. (2018). A simulation-optimization approach for scheduling in stochastic freight transportation. *Computers & Industrial Engineering*, 126:99–110.

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.

Marshall, L., Boland, N., Savelsbergh, M., and Hewitt, M. (2021). Interval-based dynamic discretization discovery for solving the continuous-time service network design problem. *Transportation science*, 55(1):29–51.

Meng, Q. and Wang, S. (2011). Liner shipping service network design with empty container repositioning. *Transportation Research Part E: Logistics and Transportation Review*, 47(5):695–708.

Moore, D. S. and McCabe, G. P. (1989). *Introduction to the practice of statistics.* WH Freeman/Times Books/Henry Holt & Co.

Müller, J. P., Elbert, R., and Emde, S. (2021a). Integrating vehicle routing into intermodal service network design with stochastic transit times. *EURO Journal on Transportation and Logistics*, 10:100046.

Müller, J. P., Elbert, R., and Emde, S. (2021b). Intermodal service network design with stochastic demand and short-term schedule modifications. *Computers & Industrial Engineering*, 159:107514.

Newbold, P., Carlson, W. L., and Thorne, B. M. (2013). *Statistics for business and economics*. Pearson.

Pedersen, M. B., Crainic, T. G., and Madsen, O. B. (2009). Models and tabu search metaheuristics for service network design with asset-balance requirements. *Transportation Science*, 43(2):158–177.

Powell, W. B. and Topaloglu, H. (2003). Stochastic programming in transportation and logistics. *Handbooks in operations research and management science*, 10:555–635.

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.

Shu, S., Xu, Z., and Baldacci, R. (2024). Incorporating holding costs in continuous-time service network design: New model, relaxation, and exact algorithm. *Transportation Science*.

Shu, S., Xu, Z., and Qi, J. (2023). Robust continuous-time service network design under travel time uncertainty. *Available at SSRN 4620781*.

Sun, Y. (2020). Green and reliable freight routing problem in the road-rail intermodal transportation network with uncertain parameters: a fuzzy goal programming approach. *Journal of Advanced Transportation*, 2020:1–21.

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.

Wang, Z. and Qi, M. (2020). Robust service network design under demand uncertainty. *Transportation Science*, 54(3):676–689.

Wieberneit, N. (2008). Service network design for freight transportation: a review. *OR spectrum*, 30(1):77–112.

Zhao, Y., Liu, R., Zhang, X., and Whiteing, A. (2018a). A chance-constrained stochastic approach to intermodal container routing problems. *PloS one*, 13(2):e0192275.

Zhao, Y., Xue, Q., Cao, Z., and Zhang, X. (2018b). A two-stage chance constrained approach with application to stochastic intermodal service network design problems. *Journal of Advanced Transportation*, 2018:1–18.

Appendices

A Scenario-based Mathematical Formulation

In the following, we report a scenario-based approximation of the stochastic two-stage model described in Section 4, where the random travel time probability distributions are approximated through a set of scenarios. Specifically, each scenario represents a possible realization of the travel time for each service leg of each potential service, and has a probability to be observed.

Let $\mathcal{S}$ be the considered set of scenarios. Each scenario $s \in \mathcal{S}$ has dimension $\left|\mathcal{A}\right|$. A probability p_s is assigned to each scenario $s \in \mathcal{S}$, with $0 \leq p_s \leq 1$ and $\sum_{s\in\mathcal{S}} p_s = 1$. Lastly, we define as τ_{as} the travel time realization for service $\sigma_a \in \Sigma$ along its service leg $a \in \mathcal{A}$ in scenario $s \in \mathcal{S}$.

We define nine sets of variables. The first stage variables are:

- $y_\sigma \in \{0,1\}$, $\sigma \in \Sigma$, indicates whether service σ is selected ($y_\sigma = 1$), or not ($y_\sigma = 0$).

The second stage variables are:

- $x^k_{as} \in \{0,1\}$, $k \in \mathcal{K}$, $a \in \mathcal{A}^{STT}$, indicates whether commodity k moves along arc a in scenario $s \in \mathcal{S}$;
- $\pi_{as} \in [0, \delta_a]$, $a \in \mathcal{A}^{STT}$, represents the postponement relative to the usual time instant at which service σ_a begins to execute (move along) arc a in scenario $s \in \mathcal{S}$;
- $\psi_{as} \geq 0$, $a \in \mathcal{A}^{STT}$, represents the time instant at which service σ_a ends to execute (move along) arc a in scenario $s \in \mathcal{S}$;
- $\varepsilon^k_{is} \geq 0$, $k \in \mathcal{K}$, $i \in \mathcal{N} \setminus \{D(k)\}$, represents the time instant at which commodity k begins its movement from terminal i in scenario $s \in \mathcal{S}$;
- $\eta^k_{is} \geq 0$, $k \in \mathcal{K}$, $i \in \mathcal{N} \setminus \{O(k)\}$, represents the time instant at which commodity k ends its movement to terminal i in scenario $s \in \mathcal{S}$;
- $r_{as} \geq 0$, $a \in \mathcal{A}$, defines the tardiness in ending to execute (move along) arc a with respect to the usual ending time in scenario $s \in \mathcal{S}$;
- $r^k_s \geq 0$, $k \in \mathcal{K}$, defines the tardiness with respect to the usual due date of commodity k at destination in scenario $s \in \mathcal{S}$;

- $z_s^k \in \{0,1\}$, $k \in \mathcal{K}$, indicates whether commodity k is outsourced from $O(k)$ to $D(k)$ in scenario $s \in \mathcal{S}$.

Mathematical formulation The problem is formulated as follows:

$$\begin{aligned}\min \sum_{\sigma\in\Sigma} f_\sigma y_\sigma + \sum_{s\in\mathcal{S}} p_s \Bigg[d^k \Big(& \sum_{k\in\mathcal{K}} \sum_{a\in\mathcal{A}^{EXT}} c_a^k x_{as}^k \\ & + \sum_{k\in\mathcal{K}} \sum_{\substack{i\in\mathcal{N}:\\ i\neq O(k),D(k)}} c_i^k \left(\varepsilon_{is}^k - \eta_{is}^k\right) \\ & + \sum_{k\in\mathcal{K}} c_{O(k)}^k \left(\varepsilon_{O(k)s}^k - o(k)\right) + \sum_{k\in\mathcal{K}} c_{D(k)}^k \left(d(k) - \eta_{D(k)s}^k + r_s^k\right) \\ & + \sum_{k\in\mathcal{K}} q^k r_s^k + \sum_{k\in\mathcal{K}} C^k z_s^k \Big) \Bigg] \\ & + \sum_{a\in\mathcal{A}} q_a r_{as} \end{aligned} \quad (29)$$

subject to:

commodity routing management

$$\sum_{a\in\mathcal{A}^{STT}:D(a)=i} x_{as}^k - \sum_{a\in\mathcal{A}^{STT}:O(a)=i} x_{as}^k = \begin{cases} -1 + z_s^k & \text{if } i = O(k), \\ 1 - z_s^k & \text{if } i = D(k), \\ 0 & \text{otherwise} \end{cases} \qquad \begin{array}{l} \forall\, k \in \mathcal{K}, \\ \forall\, i \in \mathcal{N}, \\ \forall\, s \in \mathcal{S} \end{array} \quad (30)$$

$$\sum_{k\in\mathcal{K}} d^k \left[x_{as}^k + x_{a^+s}^k\right] \leq u_a\, y_{\sigma_a} \qquad \forall\, a \in \mathcal{A}^+,\ \forall\, s \in \mathcal{S} \quad (31)$$

$$\sum_{k\in\mathcal{K}} d^k \left[x_{a^-s}^k + x_{as}^k\right] \leq u_a\, y_{\sigma_a} \qquad \forall\, a \in \mathcal{A}^-,\ \forall\, s \in \mathcal{S} \quad (32)$$

$$\sum_{k\in\mathcal{K}} d^k\, x_{as}^k \leq u_a\, y_{\sigma_a} \qquad \forall\, a \in \mathcal{A} \setminus (\mathcal{A}^+ \cup \mathcal{A}^-),\ \forall\, s \in \mathcal{S} \quad (33)$$

service time management

$$\hat{\phi}_{l_{i+1}(\sigma)} + \pi_{l_{i+1}(\sigma)s} \geq \psi_{l_i(\sigma)s} + \theta_{l_i(\sigma)}\, y_\sigma \qquad \forall\, \sigma \in \Sigma,\ i = 1, \ldots, |\mathcal{L}(\sigma)| - 1,\ \forall\, s \in \mathcal{S} \quad (34)$$

$$\psi_{as} = \hat{\phi}_a + \pi_{as} + \tau_{as} \cdot y_{\sigma_a} \qquad \forall\, a \in \mathcal{A},\ \forall\, s \in \mathcal{S} \quad (35)$$

$$\psi_{as} \leq \hat{\psi}_a + r_{as} \qquad \forall\, a \in \mathcal{A},\ \forall\, s \in \mathcal{S} \quad (36)$$

$$\pi_{a^+s} = \pi_{as} \qquad \forall\, a \in \mathcal{A}^+,\ \forall\, s \in \mathcal{S} \quad (37)$$

$$\psi_{a^+s} = \psi_{as} + T \qquad \forall\, a \in \mathcal{A}^+,\ \forall\, s \in \mathcal{S} \quad (38)$$

$$\pi_{a^-s} = \pi_{as} \qquad \forall\, a \in \mathcal{A}^-,\ \forall\, s \in \mathcal{S} \quad (39)$$

$$\psi_{a^-s} = \psi_{as} - T \qquad \forall\, a \in \mathcal{A}^-,\ \forall\, s \in \mathcal{S} \quad (40)$$

commodity time management

$$\varepsilon^k_{O(k)s} \geq o(k) \qquad \forall\, k \in \mathcal{K},\ \forall\, s \in \mathcal{S} \quad (41)$$

$$\varepsilon^k_{is} \geq \eta^k_{is} \qquad \forall\, k \in \mathcal{K},\ \forall\, i \in \mathcal{N} \setminus \{O(k), D(k)\},\ \forall\, s \in \mathcal{S} \quad (42)$$

$$\varepsilon^k_{is} \leq 2T \sum_{a \in \mathcal{A}^{STT}: O(a)=i} x^k_{as} \qquad \forall\, k \in \mathcal{K},\ \forall\, i \in \mathcal{N} \setminus \{O(k), D(k)\},\ \forall\, s \in \mathcal{S} \quad (43)$$

$$\varepsilon^k_{O(a)s} \geq \hat{\phi}_a + \pi_{as} - 2T\left(1 - x^k_{as}\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT},\ \forall\, s \in \mathcal{S} \quad (44)$$

$$\varepsilon^k_{O(a)s} \leq \hat{\phi}_a + \pi_{as} + 2T\left(1 - x^k_{a}\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT},\ \forall\, s \in \mathcal{S} \quad (45)$$

$$\eta^k_{D(a)s} \geq \psi_{as} - 2T\left(1 - x^k_{as}\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT},\ \forall\, s \in \mathcal{S} \quad (46)$$

$$\eta^k_{D(a)s} \leq \psi_{as} + 2T\left(1 - x^k_{as}\right) \qquad \forall\, k \in \mathcal{K},\ \forall\, a \in \mathcal{A}^{STT},\ \forall\, s \in \mathcal{S} \quad (47)$$

$$r^k_s \geq \eta^k_{D(k)s} - d(k) \qquad \forall\, k \in \mathcal{K},\ \forall\, s \in \mathcal{S} \quad (48)$$

$$r^k_s \leq \delta^k \qquad \forall\, k \in \mathcal{K},\ \forall\, s \in \mathcal{S} \quad (49)$$

Table 10: Sets, parameters and variables used in the mathematical models.

Sets

$\mathcal{N}$	set of physical terminals
$\mathcal{A}^{PH}$	set of physical links between terminals
Σ	set of the potential services
$\mathcal{L}(\sigma)$	set of service legs of service $\sigma \in \Sigma$
$\mathcal{A}$	set of service legs of all the potential services in Σ
$\mathcal{A}^{STT}$	set of legs of the Stochastic-Aware Service-Leg Network
$\mathcal{K}$	set of commodities
Ω	set of possible outcomes of the travel time random variable
$\mathcal{S}$	set of scenarios

Parameters

T	Schedule length
$O(k)$	origin of commodity $k \in \mathcal{K}$
$D(k)$	destination of commodity $k \in \mathcal{K}$
$o(k)$	availability date of commodity $k \in \mathcal{K}$
$d(k)$	due date of commodity $k \in \mathcal{K}$
d^k	volume of commodity $k \in \mathcal{K}$
δ^k	maximum allowed delay for commodity $k \in \mathcal{K}$
$O(\sigma)$	origin of service $\sigma \in \Sigma$
$D(\sigma)$	destination of service $\sigma \in \Sigma$
u_a	capacity of service leg $a \in \mathcal{A}$
τ_a	travel time random variable associated with service leg $a \in \mathcal{A}$, having support between τ_a^m and τ_a^M
$\tau_a(\omega)$	travel time random variable realization of service leg $a \in \mathcal{A}$
τ_{as}	observed travel time of service leg $a \in \mathcal{A}$ in scenario $s \in \mathcal{S}$
$\theta_{l_i(\sigma)}$	time spent in terminal by service $\sigma \in \Sigma$ after executing service leg $l_i(\sigma) \in \mathcal{L}(\sigma)$
$\hat{\phi}_a$	usual time instant at which service $\sigma_a \in \Sigma$ begins to execute arc $a \in \mathcal{A}$
$\hat{\psi}_a$	usual time instant at which service $\sigma_a \in \Sigma$ ends to execute arc $a \in \mathcal{A}$
δ_a	maximum allowed departure postponement of the usual time instant at which service $\sigma_a \in \Sigma$ begins to execute arc $a \in \mathcal{A}$

Continued on next page

Parameters

f_σ	activation cost of service $\sigma \in \Sigma$
c_a^k	unit transportation cost of commodity $k \in \mathcal{K}$ and arc $a \in \mathcal{A}$
c_i^k	unit holding and handling cost of commodity $k \in \mathcal{K}$ at node $i \in \mathcal{N}$
C^k	cost of outsourcing of commodity $k \in \mathcal{K}$
q^k	penalty for late arrival of commodity $k \in \mathcal{K}$ at destination
q_a	penalty for late arrival of service $\sigma_a \in \Sigma$ in ending executing its service leg $a \in \mathcal{A}$
p_s	probability of scenario $s \in \mathcal{S}$
$O(a)$	origin of arc a
$D(a)$	destination of arc a

Variables

y_σ	defines whether service $\sigma \in \Sigma$ is selected ($y_\sigma = 1$), or not ($y_\sigma = 0$)
$x_a^k(\omega)$	defines whether commodity k moves along arc a
$\pi_a(\omega)$	defines the postponement relative to the usual time instant at which service σ_a begins to execute (move along) arc a
$\psi_a(\omega)$	defines the time instant at which service σ_a ends to execute (move along) arc a
$\varepsilon_i^k(\omega)$	defines the time instant at which commodity k begins to move from terminal i
$\eta_i^k(\omega)$	defines the time instant at which commodity k ends to move to terminal i
$r_a(\omega)$	defines the tardiness in ending to execute (move along) arc a
$r^k(\omega)$	defines the tardiness of commodity k at its destination
$z^k(\omega)$	defines whether commodity k is outsourced from $O(k)$ to $D(k)$