View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by CWI's Institutional Repository

European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing, Transportation and Logistics

Minimum costs paths in intermodal transportation networks with stochastic travel times and overbookings

Bernard G. Zweers[a,b,*], Rob D. van der Mei[a,b]

[a] Centrum Wiskunde & Informatica, Stochastics, the Netherlands
[b] Vrije Universiteit Amsterdam, Department of Mathematics, the Netherlands

ARTICLE INFO

Article history:
Received 31 July 2020
Accepted 21 July 2021
Available online xxx

Keywords:
Logistics
Intermodal transport
Stochastic shortest path
Certainty equivalent
Overbookings

ABSTRACT

In intermodal transportation, it is essential to balance the trade-off between the cost and duration of a route. The duration of a path is inherently stochastic because of delays and the possibility of overbooking. We study a problem faced by a company that supports shippers with advice for the route selection. The challenge is to find Pareto-optimal solutions regarding the route's costs and the probability of arriving before a specific deadline. We show how this probability can be calculated in a network with scheduled departure times and the possibility of overbookings. To solve this problem, we give an optimal algorithm, but as its running time becomes too long for larger networks, we also develop a heuristic. The idea of this heuristic is to replace the stochastic variables by deterministic risk measures and solve the resulting deterministic optimization problem. The heuristic produces, in a fraction of the optimal algorithm's running time, solutions of which the costs are only a few percent higher than the optimal costs.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

In intermodal transport, at least two modes of transportation are used to ship goods from the origin to the destination (Macharis & Bontekoning, 2004). Two main factors influencing the route choice of a shipment in intermodal transport are the *costs* and *duration*. Usually, one is looking for the cheapest option to transport freight such that the arrival at the destination is on time. However, the travel time is to a large extent stochastic. A vehicle might be delayed or a leg of the trip could be overbooked. Therefore, it is impossible to guarantee that a shipment will arrive on time. Possible options to deal with the travel time's stochasticity are to enforce the expected arrival time of the shipment to be before its deadline, or to put a penalty on late arrival and minimize the expected costs. These methods might work fine if a shipper transports many loads, but they are troublesome if only incidentally goods are transported.

In this paper, we study a problem faced by a company that is offering a tool to shippers that ship specialized cargo on a global scale. These specialized types of cargo could, for instance, be pharmaceuticals. Medicines need to be transported in a specific temperature range because otherwise they might get ruined. The specialized goods are often shipped in cooling containers to control the temperature, but these containers only cool for a limited amount of time. When the shipment arrives after this period, it might not be at the right temperature. In these types of problems, it is more suitable to consider the *probability of on-time arrival*. Although this study originated from a collaboration with a company specialized in the shipment of pharmaceuticals, the mathematical model in this paper is so general that it can be applied by countless companies that use an intermodal network to ship their goods.

In intermodal logistics, the shipper does not operate its own fleet and uses carriers to transport the package or container. Consequently, the shipper has to obey the predefined schedule of the carrier. This schedule results in large variation in the duration of a shipment (Ziliaskopoulos & Wardell, 2000). If there is a delay in one point of the transportation chain and a departure is missed, then the arrival at the destination is often much later because there is usually some time between consecutive departures. As a result, the planned departure of the next leg might also be missed, in which the delay could further propagate through the network. In practice, it could also happen that a leg is overbooked in which the shipment can at best be transported in the next departure. The overbooking probability usually increases if one is behind its planned schedule, in which case the delay only gets worse.

In this freight network, we are interested in constructing Pareto-optimal solutions in which the probability of arrival before the deadline is compared with the costs of a route. With such a

* Corresponding author.
E-mail address: b.g.zweers@cwi.nl (B.G. Zweers).

https://doi.org/10.1016/j.ejor.2021.07.042
0377-2217/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Please cite this article as: B.G. Zweers and R.D. van der Mei, Minimum costs paths in intermodal transportation networks with stochastic travel times and overbookings, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2021.07.042

Pareto-front, the shipper can choose how much risk he or she is willing to accept to transport goods at a specific price. The number of routes grows exponentially with the number of nodes in a network and there do not exist efficient optimal algorithms. Hence, for larger-sized networks, we need to consider heuristics to find the most cost-efficient route given a certain *acceptance probability for risk*.

The contribution of this paper is three-fold. First, we model a problem faced by a company delivering an online tool for shippers as a shortest path problem with stochastic travel times and overbooking probabilities. The goal of this problem is to find the cheapest path for which the probability of arrival before a deadline is above a certain threshold. Second, we give an optimal algorithm based on dynamic programming. For this dynamic programming formulation we derive explicit formulas for the probability functions of the arrival and departure time at every node. We show that for finding dominating paths, we only need to consider a limited number of time epochs. Finally, we propose a heuristic in which the stochastic travel times are replaced by deterministic values that are a function of the risk one is willing to take. To the best of our knowledge, we are the first to use these deterministic *risk meassures* in a heuristic to solve a routing problem with stochastic travel times. After the random variables are replaced with deterministic values, an Integer Linear Program (ILP) formulation is used to solve the problem.

The remainder of this paper is organized as follows. In Section 2, a description of the relevant literature is given. After that, we describe in Section 3 in detail the problem that we are solving. This problem formulation is translated into a mathematical model in Section 4. In Section 5, the optimal algorithm and heuristic are presented to solve the model described in Section 4. The quality of this heuristic is investigated in Section 6, and finally, we conclude this paper in Section 7.

2. Literature review

In this section, we focus first on routing problems in multimodal networks. These problems are mainly deterministic and since our problem is stochastic, we also discuss the relevant literature concerning stochastic shortest paths. Finally, we review some works regarding risk measures that will be used in our heuristic.

2.1. Routing in intermodal networks

According to Chang (2008), routing problems in intermodal transport networks have three important characteristics. First, they deal with multiple objectives, such as travel time and costs. Second, the schedules and delivery times must be included to avoid a mismatch in practice. Finally, the calculation of costs is complicated because it might be dependent of the weight or volume transported. In this work, we only focus on the first two points. As in our problem setting we are concerned with transporting of a single shipment, its costs can be easily computed beforehand. Ziliaskopoulos and Wardell (2000) add that one should also account for delays at switching locations. We implicitly do that by including overbookings.

The multiple objectives can be included as a weighted sum in the objective function, or it can be decided to put one or more of the targets in the constraint. The latter can be done by constructing Pareto-optimal solutions, which is done in Cho, Kim, and Choi (2012). In this paper, Pareto-optimal solutions regarding the travel time and transportation costs are given for an international intermodal routing problem. They solve a weighted constrained shortest path problem using a dynamic programming formulation. In Gromicho, Oudshoorn, and Post (2011), a problem faced by a logistic service provider is studied in which the goal is to find the k-cheapest routes given time restrictions. This problem is also modeled as a resource constraint shortest path and is solved using a two-stage variant of Dijkstra's algorithm.

A weighted sum of the travel time and the transportation costs is minimized in Chang (2008). He enforces time-windows for the moment a path should arrive in a node and solves the resulting problem using a Lagrangean relaxation. In Yang, Low, and Tang (2011), a goal programming approach is used to minimize the weighted sum of transportation cost, transit time, and transit variability of an intermodal route. This problem is entirely deterministic because the transit variability is assumed to be a given constant. In multi-objective problems, the travel time is also sometimes assumed to be time-dependent. A multi-criteria time-dependent shortest path problem is studied in Androutsopoulos and Zografos (2009). In this problem, the scheduled departure time of an arc is fixed, and for every node, there is a strict time window for which the route should visit that node. In Ziliaskopoulos and Wardell (2000), a deterministic time-dependent shortest path problem is studied with a delay at the nodes. In Chang, Floros, and Ziliaskopoulos (2007), this problem is extended and is not only the travel time but also the costs that are associated with that path are minimized.

2.2. Stochastic shortest paths

Most classical shortest path problems with stochastic travel times deal with finding the path with the shortest expected duration (see, e.g., (Fu & Rilett, 1998; Hall, 1986; Miller-Hooks & Mahmassani, 2000)), but our problem is different in two ways. First of all, our objective is to find the cheapest path that has the largest probability of arriving on time at the destination. Second, the planned departures and the possibility of overbooking make that the arrival moment at a node is not necessarily the same as the departure time. To the best of our knowledge, we are the first to combine the two aspects described above. Nevertheless, the two problems have been studied separately, which will be discussed below.

Firstly, there is some literature on stochastic shortest path problems in which there is an arrival deadline. One concept that is applied in this context is *stochastic dominance*. A distribution stochastically dominates another distribution if, for every possible value in its domain, its cumulative distribution function is at least as large as that of the other distribution. Zhang and de Mello (2017) and Nie, Wu, and de Mello (2012) study a problem in which the goal is to find a path that minimizes the earliness and lateness and that stochastically dominates a benchmark path. In Nie et al. (2012), this problem is solved using a dynamic programming formulation, and in Zhang and de Mello (2017) a Sample Average Approximation method is proposed that can solve larger instances. In Chang, Nozick, and Turnquist (2005), an algorithm is presented to find non-dominated paths with multiple time-dependent stochastic attributes. In Cheng and Lisser (2015), a problem is studied in which the goal is to find the path for which the probability that the resource constraints are sastified is maximized, given that the costs of the route should not exceed a given threshold. Finally, in Parmentier (2019), improved bounds are presented for computing shortest paths with general non-linear and stochastic recourse constraint. These bounds can be used to speed-up an enumeration algorithm.

A shortest path problem can be modelled as a linear programming formulation. In the literature, efficient solution methods for general linear programming formulations with specific stochastic constratins have also been studied. For instance, Cheng and Lisser (2012) consider a linear program with multiple stochastic constraints in which all random variables are normally distributed. Under this assumption, the problem can be approximated

by a second-order conic program. Another example is the work of Luedtke, Ahmed, and Nemhauser (2010) in which an integer linear program formulation is proposed for the situation in which only the right-hand side of the inequality is random and has a finite distribution. Unfortunately, these techniques cannot be applied to our problem.

A branch of literature in which the stochastic travel times are combined with planned departures is the literature about *multimodal itinerary planning*. In this problem, the goal is to find a route through a public transit network for which the probability of an on time arrival is maximized (Häme & Hakula, 2013; Redmond, Campbell, & Ehmke, 2019; 2020; Zhang, Liu, Yang, & Gao, 2015). The main difference between the multimodal itinerary problem and our problem is that in the former the route can be dynamically adjusted. In a situation where a person is traveling, this assumption is realistic. However, for freight transportation it is not always applicable because a parcel cannot decide for itself that another route might be better in the new situation. Only when a planner is actively following a package and has the flexibility to change the route, the problem described in Häme and Hakula (2013) applies. In Häme and Hakula (2013), the multimodal itinerary problem is solved to optimality using a Markov decision process. In Redmond et al. (2019), a problem is studied in which the goal is to find the flight itinerary that has the largest probability of arriving on-time. In this problem, every flight has a fixed departure and if that is missed, there is no option to arrive the destination. This model is extended to a multimodal network in Redmond, Campbell, and Ehmke (2020). In Zhang et al. (2015), the different modes in the multimodal network do not have a given departure moment, but switching a mode costs extra time. The objective of their problem is to find a path for which both the total costs and the travel time are minimized but need to meet a chance constraint.

2.3. Risk measures

By a *risk measure*, we mean a function that maps a stochastic variable to a real value. A risk measure can be used to compare multiple stochastic variables. In Cominetti and Torrico (2016), multiple risk measures for shortest paths are considered. As they point out, most risk measures do not meet the *additive consistency property*. This property states that if, under a particular risk measure, one stochastic variable is preferred to another stochastic variable, then if to both these variables the same independent random variable is added, the preference relation should not change. They state that the only risk measure that has this property is the *entropic risk measure*.

An entropic risk measure that has been used before in shortest path problems is the *certainty equivalent under exponential disutility* (Jaillet, Qi, & Sim, 2016; Zhang & Tang, 2018). In Zhang and Tang (2018), this measure is applied in the context of a public transit network; their goal is to find a route that minimizes the certainty equivalent under exponential disutility, given that the arrival is before a deadline. Jaillet et al. (2016) study general routing problems for which they put constraints on the certainty equivalent under exponential disutility. They show how this framework can be applied in the case of distributionally robust optimization. As a special case, they solve a shortest path in which the least risky route with respect to the certainty equivalent under exponentially disutility has to be found such that the arrival is before the deadline.

Markowitz (1952) proposes the *expectation-variance* (EV) risk measure. In this risk measure, the weighted sum of the expectation and variance of a random variable is taken. The more weight that is given to variance, the more risk-averse the outcome is. A disadvantage of this method is that it is not monotone, which means that it could be the case that one stochastic variable is almost surely dominated by another, but that the EV risk measure prefers the latter (Cominetti & Torrico, 2016). In Hutson and Shier (2009), a more general function of the expectation and the variance is applied to a shortest path problem. The function they minimize is a sum of a convex function of the mean and a concave function of the variance.

In our heuristic, we replace a stochastic variable by a deterministic approximation, which is closely related to *Approximate Dynamic Programming* (ADP) introduced by Powell (2011). In ADP, the key idea is to use an approximation for the objective function of a dynamic program to reduce the dimensions of this problem. This technique has successfully been applied to many logistics problems, for instance see the overview papers of Powell, Simao, and Bouzaiene-Ayari (2012) and Ulmer (2017). There are two aspects that are closely related to ADP but do not apply to our problem setting. First, the stochasticity in this paper's problem only lies in the travel time which is incorporated in a constraint. In most applications of ADP, the objective function contains the uncertainty and the constraints are deterministic. Second, successful ADP applications share the flexibility to adapt the planning repeatedly during the planning process and in our problem the route is fixed before any realization of the stochastic variable is revealed.

3. Problem formulation

In our problem, two types of logistics parties are involved: the *shipper* and the *carrier*. The shipper is a company that is the supplier of the goods that need to be transported and the carrier is the company that actually transports the goods. We study a problem of a shipper who needs to ship goods through an intermodal network. The shipper is a small company that does not have its own fleet and needs to use the scheduled transport of carriers. The possible transportation modes could be airplanes, ships, trains, and trucks. The shipper has a given deadline at which the shipment should be at its destination. This deadline could be imposed by the customer that requires its shipment to arrive before a certain moment, but it could also be caused by the packaging material that is used. For instance, some packaging that is used to cool temperature sensitive material stops working after a certain amount of time. A complete route is booked at the carrier, so there is no option to change it during the transshipment. The shipper is looking for the route of the shipment that has minimal costs but also satisfies that the shipment arrives on time with a certain probability. It could be possible to allow for changes in the modalities that are used or the route of a parcel. This scenario is often referred to as *synchromodality*, but this concept is usually applied to small scaled networks and not to a global scaled networks. Moreover, a route in our problem does not only reflect the hubs through which a parcel is transshipped, but also the shipping material that is used. This shipping material is not easy to change during the transshipment.

We assume that for each leg of the transportation network the costs are known. The total costs of a route are simply the sum over all the legs it contains. The costs of a route are determined by the carrier and since only a single parcel is shipped these costs can be easily requested. Additionally, we assume that we know for each leg the distribution of its duration. It might be hard to determine the distribution of a single leg. Nonetheless, if one has data on the transportation times of an entire route then it is possible to decompose that into a distribution for a specific leg. On top of that, we assume for sake of simplicity that these stochastic variables are independent of each other. In practice, one would expect that there is a certain dependency between the transportation times. If for instance, bad weather conditions make that the shipment is delayed on one leg, it is reasonable to assume that this also influences the travel time on the subsequent leg. Although we do not have explicit dependency in our model, there is some implicit dependency within a route. Each leg has a planned departure time associated

Table 1
Overview of the notation used in this paper.

n	Destination node
c_{ij}	Costs for arc (i, j)
d_{ij}	Scheduled departure for arc (i, j)
f_{ij}	Frequency of departures on arc (i, j)
T	Deadline for arrival at node n
P	Path going from node 1 to node n
$F_P(\cdot)$	Cumulative distribution function of the arrival time at node n
ϕ_{ij}^0	Stochastic variable for the travel time on arc (i, j)
ϕ_{ij}^1	Stochastic variable for the number of departures missed before arc (i, j) is traversed if arrival at node i is before d_{ij}
ϕ_{ij}^{k+1}	Stochastic variable for the number of departures missed before arc (i, j) is traversed if arrival at node i is between $d_{ij} + (k-1)f_{ij}$ and $d_{ij} + kf_{ij}$
$\underline{\phi}$	Lower limit of domain of stochastic variable ϕ
$\overline{\phi}$	Upper limit of domain of stochastic variable ϕ
β	Risk acceptance parameter
$\mathcal{V}$	Set of all nodes
$\mathcal{A}$	Set of all arcs
$\mathcal{P}$	Set of all paths from node 1 to node n

with it, which is the first transportation possibility for that leg. Additionally, the next possible departure moments are also given. Consequently, if a shipment arrives only just after its planned departure at a hub, it can only leave that hub at the next departure time at the earliest. This later departure could have the effect that the shipment also has to wait for a more extended period at the next hub.

As the shipper is only shipping a single or a few containers through the network, we are not concerned with the capacity of a transportation link. The carrier decides if there is enough capacity left in each leg of the transportation chain to transport the shipment. However, as no-shows occur frequently, air carriers tend to overbook a flight. So it might be the case that a shipment cannot be transported on the planned time but that it will be shipped at a later moment. We assume that for each leg in the transportation network, we know the probability of how many departures are missed by a shipment. This probability depends on the planned departure. If the shipment arrives before its scheduled departure at a hub, the expected number of missed departures is less than if the shipment arrives after its planned departure. Similarly, to the distribution of the transportation time, this probability could be determined by decomposing the total transportation time for an entire path. Carriers sometimes offer different booking classes for cargo. Standard booking classes are the cheapest option, but premium booking classes have the advantage of having a higher priority when the leg is overbooked. So for the same physical route, there could be a more expensive option that has a larger probability of arriving on time.

In general, a shipper does not have a fixed value for the on-time arrival probability of the shipment. Consider a situation in which one route has a slightly larger chance of arriving too late at the destination than another route, but the costs of the former path are only a fraction of the latter. In this situation, most shippers will be inclined to take the riskier but cheaper route, but the shipper should make that decision. Therefore, the goal of this problem is to present a Pareto-front in which the probability of on-time arrival is compared with the costs of a route.

4. Mathematical model

In this section, we present a mathematical model for the problem described in Section 3. The notation that is used in this model is summarized in Table 1. We model the intermodal transportation network as a directed acyclic graph $\mathcal{G} := (\mathcal{V}, \mathcal{A})$ with node set $\mathcal{V}$ and arc set $\mathcal{A}$. Let n denote the number of nodes in a graph. There is one source node $s \in \mathcal{V}$ and a destination node $t \in \mathcal{V}$. We will number the nodes in such a way that node s gets number 1 and node t gets number n. Since the graph is acyclic it is possible to number the nodes such that there are no arcs between (i, j) if $j < i$. It could be possible that there are multiple nodes that correspond to a single hub. For instance, if the overbooking probability depends on the bookings class, one needs to make a node for every combination of hub and bookings class.

Let $\mathcal{P}$ be the set of all possible paths between node 1 and node n and we denote a single path by $P \in \mathcal{P}$. An arc (i, j) has associated costs of c_{ij}, and the total costs of all arcs in path P is denoted by $c(P)$. Additionally, an arc (i, j) has a stochastic variable ϕ_{ij}^0 for the time needed to traverse that edge and a scheduled departure time d_{ij}. For the sake of simplicity, we assume that there is a fixed time interval between the departures after d_{ij}. We call that interval between consecutive departure the *frequency* of an edge and denote it by f_{ij}. Note that our approach would also work if the departure times do not follow a specific pattern but are given beforehand. We assume that all departure times and frequencies are integers. This is realistic from a practical perspective because these times usually have a certain precision, for instance, the planned departure of a flight is usually only given with a precision of five minutes. Moreover, this assumption will be useful in computing the on-time arrival probability, as we will see in Section 5.1.

If the shipment arrives at node i before the scheduled departure time d_{ij} of arc (i, j), then the number of departures on which it cannot be shipped because there is no capacity left is the stochastic variable ϕ_{ij}^1. On the other hand, if the shipment arrives at node i between $d_{ij} + (k-1)f_{ij}$ and $d_{ij} + kf_{ij}$, then random variable ϕ_{ij}^{k+1} denotes the number of departures that is missed. In practice the expectation of ϕ_{ij}^{k+1} is larger than the expectation of ϕ_{ij}^k. The cumulative distribution function for the arrival time at node n using path P is given by $F_P(\cdot)$. On top of that, we are given a deadline $T > 0$ and a *risk acceptance threshold* $\beta \in [0, 1]$. The risk acceptance threshold is the probability for which we accept a late arrival. The larger the value of β the more risk one is willing to accept. The goal is to find a path with minimal cost such that the probability that the arrival at node n is after T is less than or equal to β. In other words, the problem we need to solve is the following:

$$\min_{P \in \mathcal{P}} \quad c(P) \tag{1}$$

$$\text{subject to:} \quad 1 - F_P(T) \leq \beta \quad \beta \in [0, 1]. \tag{2}$$

Constraint (2) is an *individual chance constraint*, and thus the problem (1)-(2) is a variation of a chance constraint problem. For a fur-

ther discussion about these types of problems we refer to Birge and Louveaux (2011).

4.1. Running example

In this paper, we use a simple network to illustrate the problem and solution methods. This network consists of four nodes and is given in Fig. 1. On each arc, two sets of three numbers are given. The first set consists of the parameters for the departure. The first value represents the planned departure time, and the second one is the frequency of the departures on that arc. Finally, the third number, p_{ij}, is the overbooking probability on that arc. For simplicity, we assume that this is the same whether the shipment arrives before or after the planned deadline. Moreover, it also does not change if more departures have already been missed. Hence, the number of missed departures ϕ_{ij}^1 is geometrically distributed with parameter $1-p_{ij}$ and 0 is in the domain of this random variable. In other words, for every arc (i, j) the probability of missing k departures is given by:

$$\mathbb{P}\left(\phi_{ij}^1 = k\right) = p_{ij}^k(1-p_{ij}) \quad k = 0, 1, \ldots$$

All departures from node 1 leave at time 0 and have no possibility of overbooking, thus the shipment will always leave that node at time 0. The second set of three parameters are concerned with the actual transportation on that arc. The first parameter represents the costs of that leg, the second the expected travel time, and the third value corresponds with the variation of the transportation time. We assume that the travel times follow the gamma distribution for two reasons. First, it is a right-skewed distribution, which reflects the situation that delays cause the mean of the distribution to be larger than its median. Second, for the gamma distribution the moment generating function and thus the certainty equivalent is well-defined. Finally, the deadline of arrival at node 4 is 20.

5. Solution method

In this section, we discuss solution methods to solve the problem presented in Section 4. We need to find a Pareto-front for the problem given in (1)-(2). A Pareto-front can be constructed by varying the values of β. The larger the value of β, the cheaper the path will be. So if we start with a value for β equal to 1, the cheapest feasible route is found. Let us denote that path by P^*. After that, we calculate for path P^* the probability that it will arrive at node n after time T. In other words, we find the value for β for which constraint (2) holds with equality for path P^*. If that value of β is found, we update the value of β in constraint (2) such that it is just slightly lower. Consequently, P^* is not longer feasible and a new path is found with higher costs and lower probability of arriving too late. We repeat the entire procedure until there is no feasible solution anymore. This technique is also known as the ε-*constraint method* (see, e.g. Ehrgott (2005)).

In the remainder of this section, we first explain how we compute the arrival distribution at the destination in the intermodal network. This method can be used in an optimal algorithm which is described in Section 5.2. However, the running time of this method grows, in the worst case, exponentially in the number of nodes in a network. Therefore, also a heuristic method is described in Section 5.3. The high level idea of this heuristic is to replace the stochastic variable by deterministic risk measures.

5.1. Computing arrival and departure distributions

Consider a path in which l nodes are visited, and denote this path by $P = (p_1, p_2, \ldots, p_l)$. The first step in solving the problem (1)-(2) is to compute $F_P(T)$. Although evaluating whether a sum of random variables is less than a certain value is in general intractable (Khachiyan, 1989), we can exploit the fact that we made the assumption that all departures are integral values. Hence, the distribution of the departures can be seen as a discrete distribution. Nonetheless, as the travel time is continuous, the arrival distribution at the next node will also be continuous. Nonetheless, this continuous distribution can be assumed to be discrete as well because as the departures only occur at discrete moments we can round up all fractional arrivals to the nearest integer.

Let us denote the probability mass function of the arrival time and the departure at node p_j by, respectively $g_j^A(\cdot)$ and $g_j^D(\cdot)$. We assume that the shipment is available at time 0 at node 1, so $g_1^A(0) = 1$. Then the arrival and departure distributions for the others nodes can be computed by the following three equations:

$$g_j^D\left(d_{j,j+1}\right) = \sum_{x=0}^{d_{j,j+1}} g_j^A(x)\mathbb{P}\left(\phi_{j,j+1}^1 = 0\right) \quad j = 1, \ldots, l-1$$

$$g_j^D\left(d_{j,j+1} + kf_{j,j+1}\right) = \sum_{x=0}^{d_{j,j+1}} g_j^A(x)\mathbb{P}\left(\phi_{j,j+1}^1 = k\right)$$
$$+ \sum_{i=0}^{k-1} \sum_{x=d_{j,j+1}+if_{j,j+1}+1}^{d_{j,j+1}+(i+1)f_{j,j+1}} g_j^A(x)\mathbb{P}\left(\phi_{j,j+1}^{i+1} = k-(i+1)\right) \quad j = 1, \ldots, l-1$$
$$k = 1, 2, \ldots$$

$$g_{j+1}^A(x) = \sum_{k=0}^{\infty} g^D(d_{j,j+1} + kf_{j,j+1})$$
$$\times \mathbb{P}\left(x - d_{j,j+1} + kf_{j,j+1} - 1 < \phi_{j,j+1}^0 \le x - d_{j,j+1} + kf_{j,j+1}\right)$$
$$j = 1, \ldots, l-1 \quad x = 0, 1, \ldots.$$

In the first equation, the probability of departing at the planned departure from a node is calculated. For this to happen, the arrival at that node should be before the planned departure and there should be no overbooking. The probability of a departure at some other time epoch is calculated in the second equation. This probability consist of two summations. The first summation is the sum of the probability of arriving before the deadline but having to miss a departure. The second summation represents the probability of arriving after the deadline times the probability of missing the correct number of departures. In the third equation, the probability of arriving at time x in node $j+1$ is calculated. This is a convolution of all possible departures at node j and the time it takes to traverse arc $(j, j+1)$ such that the arrival at node $j+1$ is between $x-1$ and x. All the convolutions required in these three equations can be efficiently computed by standard Fast Fourier Transform algorithms,see Davis (2016) for an overview. The probability mass function $g_n^A(x)$ can be used to compute the desired cumulative distribution function $F_P(T)$.

5.2. Optimal algorithm

The problem (1)-(2) is a variant of the *Resource Constrained Shortest Path Problem* (RCSPP), (see e.g. (Pugliese & Guerriero, 2013)). For this problem, two types of exact algorithms have been developed: Dynamic Programming (DP) and Lagrangian Relaxation (Lozano & Medaglia, 2013). As the Lagrangian subproblem is still hard to solve for our problem, we have to decide to develop a DP algorithm.

In the DP algorithm, we iteratively go through all the nodes and keep track of all paths entering a node. In node i, we can discard a path P_1 if there exists a path P_2 with lower costs and that *stochastically dominates* P_1. By the latter we mean that the probability of arriving at node i before a given time is for path P_2 always greater than or equal to the probability for path P_1. To formalize this concept, let $G_{P,i}^A(\cdot)$ be the cumulative distribution function of the ar-

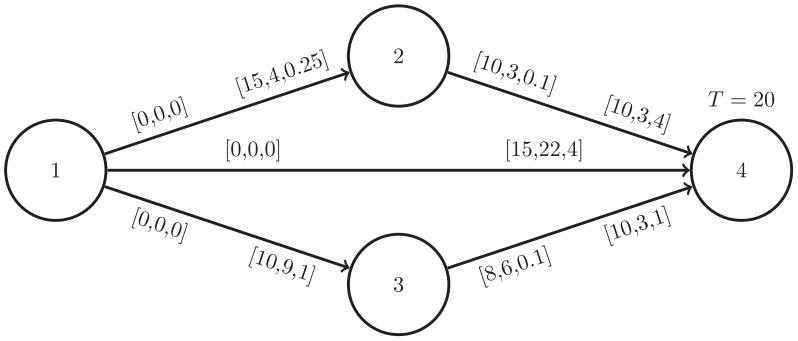

Fig. 1. Toy example of a network to illustrate the model and the solution methods.

rival time in node i in path P. So path P_2 stochastically dominates P_1 in node i if:

$$G^A_{P_2,i}(x) \geq G^A_{P_1,i}(x) \quad \forall x \geq 0. \tag{3}$$

The specifications of our problem make that it is not necessary to consider all values of x in Eq. (3). It is only possible to leave node i at a specific number of departure times, namely $D(i) := \cup_{j:(i,j)\in\mathcal{A}} \cup_{k=0}^{\infty} \{d_{ij} + kf_{ij}\}$. Every arrival in node i between two consecutive departure times can be treated as the same. Hence, we can replace the condition of Eq. (3) by:

$$G^A_{P_2,i}(x) \geq G^A_{P_1,i}(x) \quad \forall x \in D(i).$$

If path P_1 is more expensive than P_2 and is stochastically dominated by P_2, then we know that path P_1 will never be an optimal path for any acceptance threshold β. Hence, we can discard path P_1. Still, the number of paths that need to be stored can be large, especially if the set $D(i)$ is large. Therefore, we also propose a heuristic in the next section.

5.2.1. Running example (continued)

The network given in Fig. 1 has three possible routes from node 1 to node 4, namely a direct path, a route via node 2, and one via node 3. The direct path is the cheapest option with costs 15, the route via node 3 has costs 20, and the most expensive option is to ship via node 2 which has costs 25. In this example, the routes with lower costs have also a lower probability of arriving on time. The probability that the direct route arrives before 20 at node 4 is about 0.16, for the route via node 3 this probability is approximately 0.91, and finally, for the route via node 2 it is about 0.99.

In Fig. 2, the distributions of the arrival time at node 4 for the different paths are plotted. The path 1-3-4 clearly has three different peaks, corresponding to different planned departures. The first peak corresponds with the departure from node 3 at time 8, this peak is small because the probability of arriving before 8 at node 3 is not so large. The second peak resembles all departures from node 3 at time 14 and this is peak is the largest because the probability that the shipment arrives between time 8 and 14 at node 3 is large and the probability of overbooking is only 0.1. The third and smallest peak correspond with the situation of an overbooking.

Similarly, one might have expected to see also three peaks in the distribution of path 1-2-4, but in Fig. 2 there is only a single peak for this path. The reason behind this single peak is two-fold. First, the time between two departures is for edge (2,4) only 3, which is much smaller than the six time steps difference between departures for edge (3,4). Second, the variance of the travel time of edge (2,4) is also much larger than that for edge (3,4). Combining these two aspects make that the actual departure moment at node 2 has not got a big influence on the arrival moment at node 4. Hence, we do not seek three peaks for the path path 1-2-4 in Fig. 2.

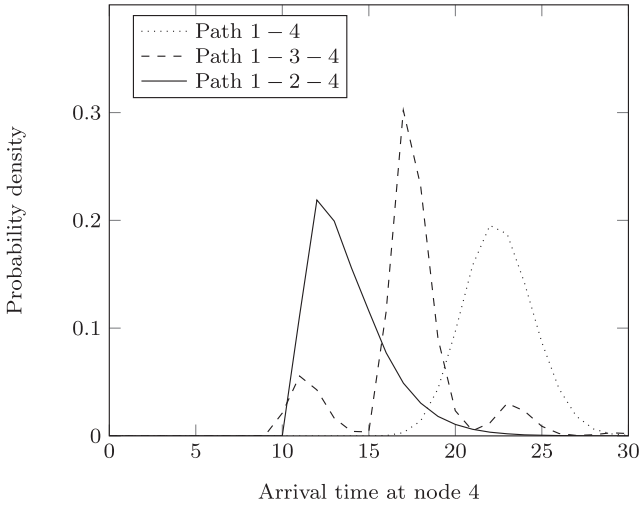

Fig. 2. Probability density function of the arrival time at node 4 for the three different paths in the network from Fig. 1.

If the three paths in this network would have been subpaths entering node 4 and there would have been an arc leaving node 4 with departure time 20, then it would not be possible to discard any of the three paths because the cheaper the path, the smaller the probability of arriving before 20 at node 4.

5.3. Risk measure heuristic

In this section, we develop a heuristic for the problem presented in Section 4. In this heuristic, all stochastic variables are replaced by a deterministic risk measure. The resulting deterministic problem is solved using an ILP. The risk measure for stochastic variable ϕ is denoted by $\rho_\alpha(\phi)$, in which α is the *risk tolerance factor*. The risk tolerance factor reflects the level of risk one is willing to take. The larger the value of α, the more risk one is accepting. If more risk is accepted than the stochastic variable ϕ is replaced by a smaller risk measure. Since if the risk measure is small than more paths result in a feasible solution. Hence, the risk measure $\rho_\alpha(\phi)$ has to be a decreasing function of α. The risk tolerance factor α can be seen as a proxy for the risk acceptance threshold β, but there is no formal relationship between the two. Moreover, the risk tolerance factor has no clear interpretation. Therefore, we start in our heuristic with small values of α, for which we will find a conservative but expensive path. If the value of α is gradually increased then all possible paths for that risk measure can be found. It is important to note that we are not interested in a Pareto-front with respect to the costs and α, but only in such a front for the costs and β. The paths constructed by this heuristic do not nec-

essarily form a set of Pareto-optimal solutions with respect to β because it could be that a path is perceived less riskier by the risk measure but that the actual probability of arriving too late is larger.

We will use two different risk measure functions: the expectation-variance (EV) function and the certainty equivalent under exponentially disutility. In the remainder of this paper, we will refer to the latter as the *certainty equivalent* (CE). We have chosen these two risk measures because they represent different levels of risk averseness. Another advantage of the EV function is that it has a clear interpretation, as we will see below, and the CE method has the benefit that it satisfies the additive consistency property, as we have seen in Section 2.

5.3.1. Expectation-variance method

The expectation-variance function for stochastic variable ϕ on the domain $[\underline{\phi}, \bar{\phi}]$ and risk tolerance factor α is defined as (Markowitz, 1952):

$$E_\alpha(\phi) = \max\left\{\underline{\phi}, \min\{\mathbb{E}(\phi) - \alpha \mathrm{Var}(\phi), \bar{\phi}\}\right\} \quad \alpha \in (-\infty, \infty).$$

The idea behind the EV method is that for positive values of α a fraction of the variance is subtracted from the expectation, and a fraction of the variance is added to the expectation for positive values of α. As we replace ϕ by $E_\alpha(\phi)$, the value $E_\alpha(\phi)$ is restricted to values that are between the lower and upper limit of ϕ. The benefit of $E_\alpha(\phi)$ is that it is easy to compute and has a relatively clear interpretation. Moreover, it can take any value in the support of ϕ and thus it can also be used for a risk-tolerant shipper.

5.3.2. Certainty equivalent method

For a random variable ϕ and risk tolerance factor α the certainty equivalent is given as follows (Jaillet et al., 2016):

$$C_\alpha(\phi) = \begin{cases} \alpha \ln \mathbb{E}\left(e^{\frac{\phi}{\alpha}}\right) & \alpha > 0 \\ \lim_{\gamma \to 0} C_\gamma(\phi) & \alpha = 0. \end{cases}$$

Using moment generating functions, $C_\alpha(\phi)$ can be easily computed for a given distribution. For instance, $C_\alpha(\phi) = \mu + \frac{\sigma^2}{2\alpha}$ if ϕ is normally distributed with mean μ and standard deviation σ or $C_\alpha(\phi) = \alpha \ln\left(\left(1 - \frac{\theta}{\alpha}\right)^{-k}\right)$ for $\alpha > \theta$ if ϕ is gamma distributed with mean $k\theta$ and standard deviation $\sqrt{k}\theta$. The function $C_\alpha(\phi)$ converges to the mean of ϕ if α goes to infinity, and it converges to the upper limit of its domain if α goes to zero (Jaillet et al., 2016). A possible downside of this approach is that the value for $C_\alpha(\phi)$ can thus never be below its mean, so it does not work very well for a risk-tolerant shipper. An advantage of the certainty equivalent is that it grows exponentially for decreasing α and that is thus it converges fast to the upper limit of the stochastic variable. Therefore, it is a suitable risk measure for a risk-averse shipper.

5.3.3. Integer linear program formulation

If each stochastic variable ϕ is replaced by a risk measure $\rho_\alpha(\phi)$, the problem becomes a deterministic optimization problem. For the moment, let us assume that the value of α is fixed, then the resulting problem can be formulated by the following ILP:

$$\min \sum_{(i,j)\in\mathcal{A}} c_{ij} x_{ij} \tag{4}$$

subject to:

$$\sum_{j:(x_{1j})\in\mathcal{A}} x_{1j} = 1 \tag{5}$$

$$\sum_{i:(i,j)\in\mathcal{A}} x_{ij} = \sum_{k:(j,k)\in\mathcal{A}} x_{jk} \quad j = 2, \ldots n-1 \tag{6}$$

$$\sum_{i:(i,n)\in\mathcal{A}} x_{in} = 1 \tag{7}$$

$$a_i \le d_{ij} + k f_{ij} z_{ijk} \quad (i,j) \in \mathcal{A} \quad k = 0, 1, \ldots, K \tag{8}$$

$$d_{ij} + \rho_\alpha\left(\phi_{ij}^0\right) + \sum_{k=0}^{K} \Big(k f_{ij} + \rho_\alpha\left(\phi_{ij}^{k+1}\right)\Big) z_{ijk} \le a_j + (1 - x_{ij})M \quad (i,j) \in \mathcal{A} \tag{9}$$

$$\sum_{k=0}^{K} z_{ijk} = 1 \quad (i,j) \in \mathcal{A} \tag{10}$$

$$a_1 = 0 \tag{11}$$

$$a_n \le T \tag{12}$$

$$a_i \ge 0 \quad i = 1, \ldots, n \tag{13}$$

$$x_{ij} \in \{0,1\} \quad (i,j) \in \mathcal{A} \tag{14}$$

$$z_{ijk} \in \{0,1\} \quad (i,j) \in \mathcal{A} \quad k = 0, 1, \ldots, K. \tag{15}$$

In this ILP, there are three types of decision variables. For each arc (i,j) we have a binary variable x_{ij} indicating whether that arc is traversed or not. Second, for every node i the decision variable a_i corresponds with the arrival moment at node i. If a path does not visit node i, the value a_i can take any value. Finally, the binary decision variable z_{ijk} is 1 if the arrival at node i is between $d_{ij} + k f_{ij}$ and $d_{ij} + (k+1) f_{ij}$ for every arc (i,j) and $k = 1, \ldots, K$. The variable z_{ij0} indicates whether the arrival at node i was before d_{ij}. The value K corresponds with the maximum number of departures that can be missed at a node and that the path still arrives before T at node n. Although, the value K is potentially different for every arc, we assume for simplicity that the maximum value over every arc is applied to all arcs.

The objective function in Eq. (4) minimizes the costs of the selected path. Constraint (5) enforces that the path leaves node 1 and constraint (7) that it arrives in node n. In constraint (6), it is ensured that if a path enters a node other than 1 or n, it also has to leave that node. The constraint (8) makes sure that the departure time $d_{ij} + k f_{ij} z_{ijk}$ from node i is after a_i, the arrival at node i. In constraint (9), it is enforced that if arc (i,j) is traversed in the path, then the arrival at node j is at least a_j. For this constraint to be valid, we need that M is a sufficiently large constant. Furthermore, the number of missed departures is given by $\sum_{k=0}^{K} \left(k f_{ij} + \rho_\alpha\left(\phi_{ij}^{k+1}\right)\right) z_{ijk}$. Finally, the term $\rho_\alpha\left(\phi_{ij}^0\right)$ represents the travel time on arc (i,j), which is independent of the actual departure time of that arc. In constraint (10), every arc (i,j) is enforced to have exactly one z_{ijk} that is equal to one. The constraints (11) and (12) ensure, respectively that the arrival at node 1 equals 0 and the arrival at node n is before the deadline T.

In practice, one might need multiple stochastic constraints. For instance, in a situation in which the freight has to be temperature regulated, the time the freight spends in a port without the right equipment to control the temperature of the shipment could also be subject to a probabilistic constraint. These types of constraints

can easily be added to the ILP above. In such a situation, one should consider carefully if also multiple values for α are needed to differentiate between the risk tolerance for different constraints.

We solve the ILP (4)-(15) using a standard ILP solver. However, as this problem is a special case of the resource constrained shortest path problem, it is NP-hard. Hence, for larger-sized instances, the computation time might be too long. Nevertheless, many exact approaches have been developed for the recourse constraint shortest path problem and they can be applied if needed. We refer to Pugliese and Guerriero (2013) for a survey on these exact algorithms.

5.3.4. Iterative procedure

In the ILP (4)-(15), it was assumed that α was fixed. As the value α has no clear interpretation, it is impossible for a practitioner to set a value of α beforehand. Nonetheless, we know that the larger the value of α the more risk is taken and the cheaper the solution will be. Therefore, the Pareto-front can be constructed in the following way: initialize $\alpha = M$ sufficiently large and $\mathcal{P} = \emptyset$. Solve for that value of α the ILP (4)-(15). Assume the path resulting path consists of l visited nodes, and let us denote this path by $P = (p_1, p_2, \ldots, p_l)$. Add this path to the set $\mathcal{P}$. After that, find the minimum value of α for which the sum of all certainty equivalents of path P is less than T. In other words, find the minimum value of α for which the solution is feasible.

A procedure to find this value for α is given in Algorithm 1. The idea of Algorithm 1 is simple, for a given value of α, the arrival time of path P at the final node n (s) is calculated. If this arrival time is lower than T, it means that the value of α for which the path arrives exactly at time T is larger. Hence, the lower bound for α, denoted by $\underline{\alpha}$ needs to be updated. If s is larger than T, then the upper bound for $\bar{\alpha}$ is updated. The function of the arrival time is discontinuous in α because if a path arrives just after d_{ij} at node i then the departure time at node i will be $d_{ij} + f_{ij}$. That is why in the while-statement in Algorithm 1, also the condition $s < T - 1$ is included. By this we ensure that in these situations we return the smaller value of α. If the value of α is found in the way described in Algorithm 1, we subtract a small value ε from it to obtain a new value for α for which the solution P is not longer feasible. We again solve the ILP (4)-(15) and repeat the procedure until the ILP (4)-(15) is infeasible.

Algorithm 1: Procedure to find the minimum α for which a given path P is feasible for ILP (4)-(15).

Input: Path $P = (p_1, p_2, \ldots, d_l)$
Initialize $s = 0$, $\underline{\alpha} = 0$ if $\rho_\alpha(\phi) = C_\alpha(\phi)$ and $\underline{\alpha} = -M$ if $\rho_\alpha(\phi) = E_\alpha(\phi)$, and $\bar{\alpha} = M$ with M being a sufficiently large number.**while** $\bar{\alpha} - \underline{\alpha} > 0.0001$ *or* $s < T - 1$ **do**
  $\alpha = \frac{\bar{\alpha} - \underline{\alpha}}{2}$
  $s = \rho_\alpha\left(\phi^0_{p_1 p_2}\right)$
  for $i = 2, \ldots, l$ **do**
    if $s \leq d_{p_i p_{i+1}}$ **then**
      $s = d_{p_i p_{i+1}} + \rho_\alpha\left(\phi^0_{p_i p_{i+1}}\right) + \rho_\alpha\left(\phi^1_{p_i p_{i+1}}\right)$
    else
      $k = \left\lceil \frac{s - d_{p_i p_{i+1}}}{f_{p_i p_{i+1}}} \right\rceil$
      $s = d_{p_i p_{i+1}} + k f_{p_i p_{i+1}} + \rho_\alpha\left(\phi^0_{p_i p_{i+1}}\right) + \rho_\alpha\left(\phi^{k+1}_{p_i p_{i+1}}\right)$
    end
  end
  if $s \leq T$ **then**
    $\bar{\alpha} = \alpha$
  else
    $\underline{\alpha} = \alpha$
  end
end
Output: α

5.3.5. Running example (continued)

We now solve our running example given in Fig. 1 with the risk measure heuristic. For this example, the solutions produced by the expectation-variance and the certainty equivalent approach are different. Recall that for α sufficiently large, the CE of a random variable is arbitrary close to its expectation. As the expectation of the travel time on the arc from node 1 to node 4 is larger than the deadline, the direct route is infeasible for the CE method. The bottom route, via node 3, is feasible for the certainty equivalent. The travel time on the first leg is 9, so the shipment is too late for the planned departure time of 8. The first departure is at 14 and then the expected number of flights being missed is the $\frac{p_{ij}}{1-p_{ij}} = \frac{1}{9}$. The expected travel time from node 3 to node 4 is 3, so the arrival time at node 4 equals $17\frac{1}{9}$. This route is feasible as long as α is roughly larger than 3.10. If α is larger than 1.44 the route via node 2 is feasible, so the heuristic using the certain equivalent will also give that route as an option.

The expectation-variance measure of a random variable takes a value smaller than its expectation as long as α is positive. Hence, if this measure is used, the direct route is found as long as $\alpha > \frac{22-20}{4} = \frac{1}{2}$. If α is just smaller than a $\frac{1}{2}$, the values taken by the EV method are close to the expectation, so the route via node 3 is also found in a similar way as for the certainty equivalent method. However, for the lowest value of α for which this route is feasible, the route via node 2 is infeasible. The EV method underestimates the risk of the path via node 3. So the risk tolerance factor for which it is still feasible is rather low.

Concluding, both the certainty equivalent method and the expectation variance method only find two of the three Pareto-optimal routes. The most risk-tolerant route is not found by the certainty equivalent method and the expectation-variance method does not return the most risk-averse route. Hence, if we combine the solutions from these two methods we do find the entire Pareto-front.

6. Numerical results

In this section, we use numerical experiments to investigate the quality of the different solution methods. We first describe in Section 6.1 how we generate random instances. Afterwards, in Section 6.2, these instances are solved using the solution methods described in Section 5 and the results are presented.

6.1. Instance generation

The different solution methods will be compared on randomly generated networks. These networks consist of n nodes with random arcs. To ensure that the graph is connected, we create an arc between every node i and $i+1$. For the other arcs, we assume that there is arc between node i and all nodes in $\left\{i+1, \ldots, \min\left\{n, i + \frac{n}{2}\right\}\right\}$ with probability $\frac{1}{2}$. This way, there could be a path from 1 to n with only one stop, but most routes will visit multiple hubs. This represents the dynamics of a intermodal network in which there is usually at least one long-haul trip from one hub to another.

Not every node in the graph corresponds with a unique physical hub. For instance, in case a shipment can be shipped using different booking classes, then every combination of hub and booking class results in a node. A higher booking class could potentially have a lower overbooking probalility, but in this paper, for the sake of simplicity, we just consider nodes and do not make a differentiation between booking classes.

The costs of an arc (i, j) are randomly generated as follows: $c_{ij} = (j - i) * \text{Uniform}(1, 4)$. This ensures that the expected costs of an outgoing arc of i are larger if j the length of an arc is also larger. The duration of an arc is gamma distributed and the mean duration of an arc (i, j) is randomly uniform between 3 and 10, and thus, it is independent of i and j. The standard deviation is randomly uniform between 0.2 and 3.

We assume that the number of departures missed because of overbooking is geometrically distributed. The parameter of this distribution is 0.9 if the shipment arrives before its planned departure at a node and 0.75 if it arrives after its planned departure. Hence, we assume that all Φ_{ij}^k are the same for every $k \geq 2$, and thus we will use Φ_{ij}^2 for every Φ_{ij}^k for $k \geq 2$. The value 0 is included in the domain of the geometric distribution to make sure that is possible to miss no departures. We include the fact that a shipment is always shipped with the fifth departure, so it cannot miss more than four departures. Hence, the probability distributions for ϕ_{ij}^1 and ϕ_{ij}^2 are as follows for every arc (i, j):

$$\mathbb{P}\left(\phi_{ij}^1 = k\right) = \begin{cases} 0.1^k 0.9 & k = 0, 1, 2, 3 \\ 0.1^4 & k = 4 \end{cases}$$

$$\mathbb{P}\left(\phi_{ij}^2 = k\right) = \begin{cases} 0.25^k 0.75 & k = 0, 1, 2, 3 \\ 0.25^4 & k = 4. \end{cases}$$

For the scheduled departure on an arc (i, j), we calculate the shortest path with respect to the mean duration to node i and add a discrete random uniform number between 3 and 13 to it. The frequency of an arc is also a discrete number that is randomly generated from the uniform distribution between 4 and 20. We assume that every arc leaving node 1 has planned departure time 0 and is never overbooked, so it is always possible to leave this node at that time. For the final deadline, the shortest path to the destination with respect to the mean duration of the arcs is calculated. It is very unlikely that for a reasonable risk acceptance threshold, there exists a feasible solution for this deadline. Hence, we generate multiple instances with different *deadline factors* (T_f). We multiply the final deadline with this deadline factor to obtain deadlines that are less restrictive. In our instances, we have chosen 2, 3, 4, and 5 as deadline factor. The deadline factor reflects the behavior of a shipper that if there does not exist a route that gives a desired on-time arrival probability that then the deadline is set later.

6.2. Comparison solution methods

In this section, we compare the optimal solution with the heuristic method using 100 instances of 50 nodes that are randomly generated as described in Section 6.1. We have chosen for this network size because its a realistic size especially if there are two to four different bookings classes. Moreover, the optimal algorithm can still solve it in a few minutes, but for larger instances the running time of the optimal algorithm becomes problematic. For example, for a network consisting of 250 nodes, the average running is about an hour. The heuristic can still solve the problem for this size of networks in a few seconds.

To compare the quality of the heuristic paths with the optimal, we compute the actual probability of arriving late at the destination for every heuristic path. If it turns out that a path that was conceived as less risky by the heuristic is actually riskier and more expensive than another heuristic path, then it is removed from the set of heuristic paths. As the running time of the heuristic is relatively short, it is also possible to compute the solution of both the certainty equivalent and expected variance heuristic and take the best solution of the two. So in this section we will compare the EV heuristic, the CE heuristic, and the combination of these two (EV+CE) with the optimal solution. We assess the quality of the heuristic at two levels. First, we check if for a certain value of β a feasible solution is found. A good heuristic should have a high probability of finding a path if there exists a solution. The second level is the cost of a route produced by the heuristic. If the solution heuristic returns a path, then its costs should ideally be close to the optimal costs.

Table 2
Number of instances for which for different solution methods and acceptance thresholds feasible paths are found.

T_f	Method	β					
		0.5	0.25	0.1	0.05	0.025	0.01
2	Opt	87	45	1	0	0	0
	EV	49	26	1	0	0	0
	CE	9	30	1	0	0	0
	EV+CE	67	35	1	0	0	0
3	Opt	100	100	87	71	57	21
	EV	85	76	49	31	20	6
	CE	96	94	73	57	46	19
	EV+CE	97	96	77	60	48	19
4	Opt	100	100	100	100	100	85
	EV	92	87	79	68	60	33
	CE	95	93	90	86	84	66
	EV+CE	96	95	92	86	86	69
5	Opt	100	100	100	100	100	100
	EV	93	90	84	78	68	51
	CE	96	93	91	91	89	86
	EV+CE	97	95	93	93	91	87

Table 3
Average percentage difference in costs for a method with the optimal solution if method has found a solution for that instance.

T_f	Method	β					
		0.5	0.25	0.1	0.05	0.025	0.01
2	EV	4.4%	3.4%	0.0%	0.0%	0.0%	0.0%
	CE	15.6%	5.3%	0.0%	0.0%	0.0%	0.0%
	EV+CE	5.4%	2.5%	0.0%	0.0%	0.0%	0.0%
3	EV	2.1%	2.7%	5.6%	3.7%	1.4%	0.0%
	CE	10.0%	5.0%	2.7%	2.7%	2.2%	0.0%
	EV+CE	3.1%	2.1%	1.5%	2.2%	2.1%	0.0%
4	EV	1.0%	5.2%	9.6%	12.1%	11.7%	5.9%
	CE	4.2%	2.4%	1.6%	3.7%	3.5%	3.7%
	EV+CE	1.5%	0.7%	1.0%	1.8%	2.5%	3.3%
5	EV	0.8%	2.2%	4.3%	8.0%	9.4%	6.6%
	CE	2.5%	2.5%	2.4%	1.9%	0.9%	1.1%
	EV+CE	0.5%	2.0%	1.5%	1.2%	0.8%	0.7%

In Table 2, for different values of β and T_f it is shown for how many instances a solution is found by the different solution methods. By increasing the value of β or T_f, we see that the number of feasible solutions increases. Moreover, it can be concluded that the number of instances for which a feasible path is found by the EV and CE heuristic on their own is for certain combinations of β and T_f much lower than the instances for which the optimal solution gives a solution. However, by combining the two heuristics the number of feasible paths is much higher. Hence, we can conclude that the two different risk measures produce sufficiently different solutions. It should also be noted that the CE heuristic returns for more instances a feasible path, which confirms the idea that it is more risk-averse than the EV heuristic.

In Table 3, the costs of the solutions returned by the three heuristics is compared with the optimal costs. To make a fair comparison, we condition the costs for the optimal solution for all three methods only on the instances for which that heuristic finds a feasible solution. Hence, the number of instances for which we compute the optimality gap is the number given in Table 2. This could lead to outcomes that at first sight might be unexpected. For instance, for $T_f = 2$ and $\beta = 0.5$ the optimality gap for the EV heuristic is smaller than optimality gap for the combination of the

Table 4
Average hypervolume indicator over all instances for the different solution methods and deadline factors.

	T_f			
	2	3	4	5
OPT	0.35	0.59	0.67	0.69
EV	0.22	0.47	0.59	0.63
CE	0.17	0.52	0.61	0.65
EV+CE	0.28	0.56	0.64	0.67

Table 5
Total number of paths found by the different methods.

T_f	Opt		EV		CE		EV+CE	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
2	2.0	1.2	0.9	0.9	0.2	0.4	0.9	0.9
3	4.5	2.1	2.4	1.5	2.6	1.4	2.9	1.5
4	6.2	2.4	3.1	1.4	2.6	1.2	4.1	1.7
5	6.1	2.5	3.0	1.3	2.9	1.3	3.9	1.6

EV and CE heuristic. The solution returned by the combination of the EV and CE heuristic is at least as good as the solution from the EV heuristic. However, the combination heuristic finds solutions to more instances and these instances are likely to be harder. Hence, it is correct that the optimality gap for the combination heuristic is larger than for the EV heuristic. Nevertheless, in general the solutions that are obtained by combining are much better than those of the single heuristics and have an optimality gap of about 2%.

As we concluded before, the solutions from the CE heuristic are more conservative than those of the EV heuristic and this effect can also be seen in Table 3. For $\beta = 0.5$, the solutions produced by the CE heuristic have a much larger gap with the optimal solutions than the solutions from the EV heuristic. This is caused by the fact that the CE heuristic often only produces solutions that have a much lower acceptance threshold than 0.5. For instance, if there is a cheap route for which the probability of arriving too late is 0.4 it is more likely to be found by the EV heuristic than by the CE heuristic. Nonetheless, if β decreases, the relative quality of the solutions by the CE method improves compared with the EV heuristic.

In Table 2, it is shown how many feasible paths were found, and the costs of these solutions are investigated in Table 3. By looking at the *hypervolume indicator*, we can simultaneously look at the feasibility and solution quality of paths. For a detailed description how to compute the hypervolume indicator we refer to Lacour, Klamroth, and Fonseca (2017) and Jaszkiewicz (2018). The idea behind the hypervolume indicator is to compute the area above the Pareto front of a set of solutions produced by a method. The larger this area, the lower the objective function, and thus, the better the solution method. The area above the Pareto curve is unbounded so we need to derive an upper bound for the costs of the solution. It is not trivial to compute such an upper bound but for all instances that we solve and all solution methods we can check that the costs never exceed 200, so we use that as an upper bound. For every value of $\beta \in \{0.01, 0.02, \ldots, 0.99\}$ we solve all instances with all solution methods. If a solution method produces a feasible solution for a value of β, we subtract the cost of that solution from 200. If this value is compute for all $\beta \in \{0.01, 0.02, \ldots, 0.99\}$ and we sum over these values, then an approximation for the area above the Pareto-front is obtained. To give a more intuitive meaning to this area, we normalize it such that it is always between 0 and 1. To do so, it is important to realize that the costs of a path is always positive, so the area above the Pareto-front is at most 200 times the number of values of β that are considered, i.e., $200 \cdot 99 = 19,800$. Hence, if the area above the curve Pareto front is divided by 19,800, the hypervolume indicator is obtained which is always a value between 0 and 1.

In Table 4, the average hypervolume indicator for the different solution methods and deadline factors over all 100 instances is shown. We see that the EV heuristic outperforms the CE heuristic if T_f equals 2, but that for larger values of T_f the CE heuristic is better. This observation supports the claim that the EV heuristic is more suitable for conservative planners. If the deadline is tight its performance is better than the CE heuristic but if more transportation time is allowed than the CE heuristic produces better solutions. Moreover, the combination of the two heuristic is much closer to the optimal solution than the two heuristic independent. Finally, the larger the deadline factor, the closer the costs of the heuristic's solutions are to the optimal costs. An explanation for this observation is that the larger the deadline factor, the more paths are feasible and thus, finding a feasible solution that is close to the optimal solution is easier.

Finally, we look at the number of solutions produced by the different methods. The objectives of finding a path that arrives on time with sufficient probability and that has minimal costs are not conflicting if only a single solution is returned. We again solved the hundred instances for every $\beta \in \{0.01, 0.02, \ldots, 0.98, 0.99\}$. In Table 5, we give the average number of paths found for an instance and its standard deviation. We see that the average number of optimal paths for an instance is about 4 to 6. The number of paths returned by the two heuristics is lower with a value between 2 and 3, but, again by combining these two heuristics, the number of paths that is found is increased to an average value between 3 and 4. For T_f equals two, the number of average paths that is found is lower because there are fewer feasible paths. Furthermore, as we noted before the CE heuristic is more conservative than the EV heuristic and thus, the number of paths returned by the CE heuristic is also lower than for the EV heuristic.

7. Conclusion

In this paper, we studied an intermodal routing problem inspired by a company that is offering shippers a tool to find the best route for their shipment. In deciding what the best route is, a trade-off has to be made between the shipment costs and the probability of arriving before a deadline at the destination. A distinct feature of our model is that it includes two types of stochasticity. First of all, the travel times between two nodes are stochastic. Moreover, we also added the possibility of overbooking as this happens on a regular basis in practice and has a major influence on the arrival time of a shipment. All legs in our transportation network have a planned departure time and if the shipment arrives after the planned departure time at the origin of this leg, then the probability of overbooking gets even larger.

We have shown how to calculate, for a given path, the probability of arriving on time at the destination. As it could be hard to find dominating paths in our model, an optimal algorithm might need too much computation time for practical use. Therefore, we have proposed a heuristic in which the stochastic variables are replaced by two different deterministic risk-measures. The resulting problem is a RCSPP that we solve using an ILP-formulation. This heuristic produces solutions that are close to the optimal solution and its running time is significantly smaller than the optimal algorithm.

The heuristic can solve networks with 250 nodes in a few seconds and the optimal algorithm needs for these networks an hour. For even larger networks, also the running time of the heuristic might become too long to use in practice because multiple ILPs need to be solved. So an interesting direction for further research would be to find a heuristic that can solve larger instances. One way to do this could be the use of exact algorithms tailored for

the resource constraint shortest path problem, instead of using the standard ILP implementation. Another way to improve the running time of our heuristic could be to use a heuristic to solve the RCSPP. The solution quality of the heuristic could potentially be improved by another risk measure. It would be interesting to investigate risk measures that are more risk averse than the EV heuristic but less risk averse that the CE heuristic.

The current problem formulation assumes that the complete path is booked in advance. If one would allow for adjustments in the path after a missed departure, the described solution methods could still be used but their quality might be worse. At every node of the path we could recalculate the best path for the remaining graph. However, the flexibility of re-optimizing might result in a different path from the beginning than the proposed path if the entire trip is booked in advance. For instance, it could be that from a specific node there is one path to the destination that is good if the shipment arrives relatively early but bad if the arrival at that node is relatively late. If there is another path for which it is the other way around, then the performance of both paths might be sub-optimal if the path needs to be booked in advance. Nevertheless, if we allow for re-optimizing these paths might have a good quality. Hence, finding good solutions for a synchromodal scenario in which the route can be changed after a disruption needs to be investigated in further research.

In our current model, the planned departure time for a leg is assumed to be given. An extension to this model would be to decide on the first possible departure. This departure should then be in a set of possible departure times. This model would better reflect reality but it also adds extra complexity to the problem. If one decides to plan the first departure later, then the probability of arriving before that time increases and thus the probability of overbooking decreases. On the other hand, there is no option to leave earlier than the first departure, so the probability of arriving on time at the destination could also decrease.

Acknowledgments

This work was supported by the Lane Analysis and Route Advisor project, funded by the Dutch Institute for Advanced Logistics (DINALOG)(project number 2018-2-171TKI).

References

Androutsopoulos, K. N., & Zografos, K. G. (2009). Solving the multi-criteria time-dependent routing and scheduling problem in a multimodal fixed scheduled network. *European Journal of Operational Research, 192*, 18–28.
Birge, J., & Louveaux, F. (2011). Introduction to stochastic programming. *Springer Series in Operations Research and Financial Engineering*. Springer.
Chang, E., Floros, E., & Ziliaskopoulos, A. (2007). An intermodal time-dependent minimum cost path algorithm. In V. Zeimpekis, C. D. Tarantilis, G. Giaglis, & I. Minis (Eds.), *Dynamic fleet management*. In *Operations Research/Computer Science Interfaces Series: 38*. Springer.
Chang, T. (2008). Best routes selection in international intermodal networks. *Computers and Operations Research, 35*, 2877–2891.
Chang, T., Nozick, L., & Turnquist, M. (2005). Multiobjective path finding in stochastic dynamic networks, with application to routing hazardous materials shipments. *Transportation Science, 39*(3), 383–399.
Cheng, J., & Lisser, A. (2012). A second-order cone programming approach for linear programs with joint probabilistic constraints. *Operations Research Letters, 40*, 325–328.
Cheng, J., & Lisser, A. (2015). Maximum probability shortest path problem. *Discrete Applied Mathematics, 192*, 40–48.
Cho, J., Kim, H., & Choi, H. (2012). An intermodal transport network planning algorithm using dynamic programming–a case study: From Busan to Rotterdam in intermodal freight routing. *Applied Intelligence, 36*, 529–541.
Cominetti, R., & Torrico, A. (2016). Additive consistency of risk measures and its application to risk-averse routing in networks. *Mathematics of Operations Research, 41*(4), 1510–1521.
Davis, J. (2016). *Methods of applied mathematics with a software overview* (2nd). Birhaeuser.
Ehrgott, M. (2005). *Multicriteria optimization* (2nd). Springer.
Fu, L., & Rilett, L. (1998). Expected shortest path in dynamic and stochastic traffic networks. *Transportation Research Part B, 32*(7), 499–516.
Gromicho, J., Oudshoorn, E., & Post, G. (2011). Generating price-effective intermodal routes. *Statistica Neerlandica, 65*(4), 432–445.
Hall, R. (1986). The fastest path through a network with random time-dependent travel times. *Transportation Science, 20*(3), 182–188.
Hutson, K., & Shier, D. (2009). Extended dominance and a stochastic shortest path problem. *Computers and Operations Research, 36*, 584–596.
Häme, L., & Hakula, H. (2013). Dynamic journeying under uncertainty. *European Journal of Operational Research, 225*, 455–471.
Jaillet, P., Qi, J., & Sim, M. (2016). Routing optimization under uncertainty. *Operations Research, 64*(1), 186–200.
Jaszkiewicz, A. (2018). Improved quick hypervolume algorithm. *Computers & Operations Research, 90*, 72–83.
Khachiyan, L. (1989). The problem of calculating the volume of a polyhedron is enumerably hard.. *Russian Mathematical Surveys, 44*(3), 199–200.
Lacour, R., Klamroth, K., & Fonseca, C. (2017). A box decomposition algorithm to compute the hypervolume indicator. *Computers & Operations Research, 79*, 347–360.
Lozano, L., & Medaglia, A. (2013). On an exact method for the constrained shortest path problem. *Computers & Operations Research, 40*, 378–384.
Luedtke, J., Ahmed, S., & Nemhauser, G. (2010). An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming Series A, 122*, 247–272.
Macharis, C., & Bontekoning, Y. (2004). Opportunities for or in intermodal freight transport research: A review. *European Journal of Operational Research, 153*, 400–416.
Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77–91.
Miller-Hooks, E., & Mahmassani, H. (2000). Least expected time paths in stochastic, time-varying transportation networks. *Transportation Science, 34*(2), 198–215.
Nie, Y., Wu, X., & de Mello, T. H. (2012). Optimal path problems with second-order stochastic dominance constraints. *Networks and Spatial Economics, 12*, 561–587.
Parmentier, A. (2019). Algorithms for non-linear and stochastic resource constrained shortest path. *Mathematical Methods of Operations Research, 89*, 281–317.
Powell, W. (2011). *Approximate dynamic programming: Solving the curses of dimensionality* (2nd). Wiley.
Powell, W., Simao, H., & Bouzaiene-Ayari, B. (2012). Approximate dynamic programming in transportation and logistics: A unified framework. *EURO Journal on Transportation and Logistics, 1*, 237–284.
Pugliese, L. D. P., & Guerriero, F. (2013). A survey of resource constrained shortest path problems: Exact solution approaches. *Networks, 62*(3), 183–200.
Redmond, M., Campbell, A. M., & Ehmke, J. (2019). The most reliable flight itinerary problem. *Networks, 73*, 325–343.
Redmond, M., Campbell, A. M., & Ehmke, J. (2020). Data-driven planning of reliable itineraries in multi-modal transit networks. *Public Transport, 12*, 171–205.
Ulmer, M. (2017). Approximate dynamic programming for dynamic vehicle routing. *Operations Research/Computer Science Interface Series*: 61. Springer.
Yang, X., Low, J., & Tang, L. (2011). Analysis of intermodal freight from china to indian ocean: A goal programming approach. *Journal of Transport Geography, 19*, 515–527.
Zhang, L., & de Mello, T. H. (2017). An optimal path model for the risk-averse traveler. *Transportation Science, 51*(2), 518–535.
Zhang, Y., Liu, P., Yang, L., & Gao, Y. (2015). A bi-objective model for uncertain multi--modal shortest path problems. *Journal of Uncertainty Analysis and Applications, 3*(8).
Zhang, Y., & Tang, J. (2018). Itinerary planning with time budget for risk-averse travelers. *European Journal of Operational Research, 267*, 288–303.
Ziliaskopoulos, A., & Wardell, W. (2000). An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operational Research, 125*(3), 486–502.