CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

Scheduled Service Network Design with Resource Management for Two-Tier Multimodal City Logistics

Pirmin Fontaine
Teodor Gabriel Crainic
Ola Jabali
Walter Rei

June 2019

CIRRELT-2019-22

Scheduled Service Network Design with Resource Management for Two-Tier Multimodal City Logistics[ł]

Pirmin Fontaine[1], Teodor Gabriel Crainic[2,*], Ola Jabali[3], Walter Rei[2]

[1] Technical University of Munich, School of Management, Munich, Germany

[2] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Management and Technology, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Canada H3C 3P8

[3] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milano 20133, Italy

Abstract. We address the tactical-planning problem for an extended two-tiered City Logistics system. This more realistic problem setting, compared to the literature, integrates inbound and outbound demands, different transportation modes combining traditional, road-based, carriers and massive-ow modes and vehicles such as light and regular rail, assignment of customers to consolidation distribution centers and satellites, and the management of a number of major resources, such as the multiple facets of satellite capacity and the structure, allocation, and size of the heterogeneous fleets. We propose a scheduled service network design formulation for the tactical planning of such extended systems, as well as an efficient Benders decomposition algorithm, which includes a tailored partial decomposition technique for deterministic mixed-integer linear-programming formulations. The results of extensive numerical experiments show the efficiency of the proposed solution method, as well as the benefits of integrating several demand types and multimodal transportation networks into a single formulation.

Keywords: Two-tier city logistics, service network design, Benders decomposition, tactical planning.

Acknowledgements. While working on the project, the first author was postdoctoral fellow at the School of Management, Université du Québec à Montréal and member of the Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), while the second author was Adjunct Professor, Department of Computer Science and Operations Research, Université de Montréal. We gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery and Acceleration Grant programs. We also gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants, and their new university researchers start up program. Thanks are also due to Calcul Québec and Compute Canada for providing the authors access to their high-performance computing infrastructure.

[ł] Revised version of the CIRRELT-2017-27

---

* Corresponding author: teodorgabriel.crainic@cirrelt.net

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2019

1 Introduction

In its most fundamental meaning, *City Logistics* aims to reduce the externalities and nuisances, e.g., emissions and, more generally, the environmental footprint, associated to the transportation of freight within urban areas, while sustaining the social and economic development of the organizations and cities involved (Benjelloun and Crainic, 2009; Bektaş et al., 2017; Crainic, 2008; Savelsbergh and Van Woensel, 2016; Taniguchi et al., 2001; Taniguchi, 2014). City Logistics encompasses several dimensions, and may target demand estimation (e.g., Holguín-Veras et al., 2014) or, more often, the design and organization of the supply activities servicing it. Studies on City Logistics may focus on a single organization (e.g., Winkenbach et al., 2016; Huang et al., 2019) or on several organizations interacting in a city (e.g., Crainic and Montreuil, 2016). A large part of the City Logistics literature concerns developing models and optimization procedures to strategic (e.g., Gianessi et al., 2016), tactical (e.g., Crainic et al., 2009) or operational (e.g., Barceló et al., 2007) planning problems.

Our work addresses the tactical planning of the supply side of *Two-Tier City Logistics*, *2T-CL*, systems for medium and large urban areas. 2T-CL systems are typically composed of two types of facilities, inbound freight being first sorted and consolidated at large *City (Urban) Distribution Centers* (*CDC*s) located on the outskirts of the city, before being transported to *satellites*, secondary, transdock-type facilities, for transfer to smaller, generally environment-friendly vehicles, for distribution to the final destinations. Two-tier systems have been proposed both for single organizations and for multiple organizations (e.g., carriers and other service providers) operating in (parts of) cities under some form of cooperation and resource sharing (e.g., Crainic et al., 2004; Morana et al., 2014). 2T-CL are consolidation-based transportation systems involving multiple resources in complex interactions, and thus require advanced planning methods, particularly at the tactical level where system-wide decisions are made for a medium-term planning horizon. Tactical planning aims to determine the service network and schedule of a consolidation-based transportation system, together with the distribution of demand flows using the selected services and the allocation and management of resources required to support them, to ensure the system is efficient and profitable, while achieving the service-quality goals set by the carrier and customers (Crainic and Laporte, 1997; Crainic, 2003; Crainic and Kim, 2007). The resulting tactical plan guides operations and provides an important evaluation instrument for strategic planning. *Scheduled Service Network Design* (*SSND*) is the generally adopted methodology to address tactical planning issues in consolidation-based transport systems (Crainic, 2000; Crainic and Kim, 2007). SSND models are generally built on time-space networks, each potential service being defined as a path (with particular operational characteristics) from its origin-terminal-initial-departure-time node to its destination-terminal-final-arrival-time node, through a number of intermediate nodes (and holding arcs at those). SSND problems are generally formulated as mixed-integer linear programming, MILP, models, which select services from a set of potential ones and route the demand flows to minimize the total generalized cost of the system (see, e.g., Crainic et al., 2014, 2017; Erera et al., 2013, for recent contributions to the SSND literature).

Crainic et al. (2009) proposed a SSND modeling framework in their pioneering work on planning 2T-CL systems, which combines particular settings of service network design for the first tier and vehicle routing for the second. The authors introduced a meta-heuristic based on decomposing the problem along the tiers, but without actually implementing it. They also observed that for longer-term planning issues, e.g., evaluation of strategic alternatives and season-long tactical plans, the second-tier routing problem could be approximated and added to the first-tier SSND formulation. The problem addressed in Crainic et al. (2009) considered a basic problem setting, however, with a single road-based transportation mode at each tier and inbound demand only, each individual inbound demand being pre-assigned to a CDC. Most publications that followed investigated the same or simpler settings, either as two-echelon vehicle or location-routing problems (e.g., Perboli et al., 2011; Hemmelmayr et al., 2012; Gianessi et al., 2016; Boccia et al., 2017) or as SSND (e.g., Crainic and Sgalambro, 2014; Crainic et al., 2016a). The Crainic et al. (2009) model is still the most comprehensive formulation for tactical planning of 2T-CL systems. Yet, neither it nor the rest of the literature accounts for several city logistics elements and concerns that are strongly emerging in the scientific literature and practice, including the utilization of public transportation modes such as light rail and buses (e.g., Arvidsson and Browne, 2013; Ghilas et al., 2013; Lindholm and Behrends, 2012; Masson et al., 2017; Savelsbergh and Van Woensel, 2016; Freemark, 2011; Danard and Janin, 2016; Riemann, 2019; Brandt, 2018), and the simultaneous consideration of inbound and outbound demand (e.g., Nguyen et al., 2017).

Our objective is to contribute to fill up these gaps in knowledge and decision-support technology. We address a more realistic problem setting in both demand and supply. On the former side, the model we propose simultaneously optimizes the transportation of inbound and outbound origin-to-destination demands by the selected services, and decides on the CDC and satellite to use for each inbound and outbound demand. On the supply side, we explicitly consider a multimodal setting, combining traditional, road-based, modes and massive-flow modes and vehicles such as light and regular rail. We also model the resources of the system, in particular, the structure, allocation, and size of the heterogeneous fleets, as well as the three-dimentional satellite capacity (in terms of the total volume of goods it may handle at any given time period, and the total and mode-specific vehicles it may accommodate simultaneously).

We propose a new SSND model for this extended problem setting, as well as an efficient Benders decomposition algorithm capable of solving quite large instances of this complex problem. The performance of our algorithm is enhanced by the development of a set of specialized valid inequalities and an innovative partial decomposition strategy that is based on the use of aggregation techniques for deterministic problems. The numerical results show significant computational benefits when using our proposed method compared to a well-known commercial solver. The experiments we report also show that considering different transportation modes and combining inbound and outbound flows, while accounting for resources, yields significant benefits, and should be considered when City Logistics systems are contemplated.

To sum up, the contributions of this paper are: (1) Defining a new, more realistic problem setting, compared to the literature, for two-tier City Logistics systems that integrates inbound and outbound demands, different transportation modes, assignment of demand flows to CDCs and satellites, and resource management; (2) Proposing a Scheduled Service Network Design formulation for the tactical planning of such extended 2T-CL settings; (3) Developing an efficient Benders decomposition algorithm for the proposed SSND, which includes a tailored partial decomposition technique for deterministic MILP formulations; According to our best knowledge, this is the first exact solution method proposed for a SSND tactical-planning model for City Logistics; (4) Showing through extensive numerical experiment, the efficiency of the proposed solution method, as well as the benefits of integrating several demand types and multimodal transportation networks.

The remainder of the paper is structured as follows. We define the problem setting in Section 2, and relate it to the existing literature in in Section 3. The SSND model is introduced in Section 4. Section 5 is dedicated to the presentation of the Benders decomposition algorithm and the improvement strategies we developed. Numerical results and managerial insights are given in Section 6. The paper concludes with a summary of the work and an outlook for future research directions.

2 Problem Setting

We first introduce the *Two-Tier Multimodal City Logistics* (*2TM-CL*) system we address, its basic components, and operation principles. We emphasize the new contributions with respect to the literature, which is briefly reviewed in Section 3. We complete this section with a discussion of tactical planning for the 2TM-CL system we address in this paper, which is modelled as a SSND in Section 4.

2.1 The 2TM-CL System

We address the City Logistis tactical planning problem in a 2TM-CL setting, which generalizes and greatly expands the structures and settings considered previously in the literature (Crainic et al., 2009; Gianessi et al., 2016; Bektaş et al., 2017; Taniguchi et al., 2014; Savelsbergh and Van Woensel, 2016; Mancini et al., 2014; Bektaş et al., 2017). We continue to assume, however, that the CL system is planned and managed by a single entity manager/decision-maker, even though resources may be provided and operated by several private and public stakeholders involved in some form of cost/profit/risk-sharing collaboration (see, e.g., Rusish, 2017, for a taxonomy of collaborative logistics).

The 2TM-CL system, illustrated in Figure 1, consists of two layers of terminals, also called *tiers* following the vocabulary proposed by Crainic et al. (2009), linked by transportation means of various modes, the latter greatly expanding previous problem settings. The goal is to satisfy demand, that is, to deliver goods from origins to destinations, on time, economically in monetary terms, and efficiently from the societal point of view of the impact on the city. This impact accounts for the street or neighborhood characteristics, e.g., touristic, residential, social (schools, hospitals, .etc.), leisure, etc., and can be defined in terms of emissions, noise, visual degradation, contribution to congestion, and so on. The particular measures considered, as well as their estimation processes, are, of course, application specific. For tactical-planning purposes, and similarly to previous studies (e.g., Crainic et al., 2009), economic and impact measures are combined into a *transportation* or *city-infrastructure-utilization cost* associated to the network representation used in the model (Section 4).

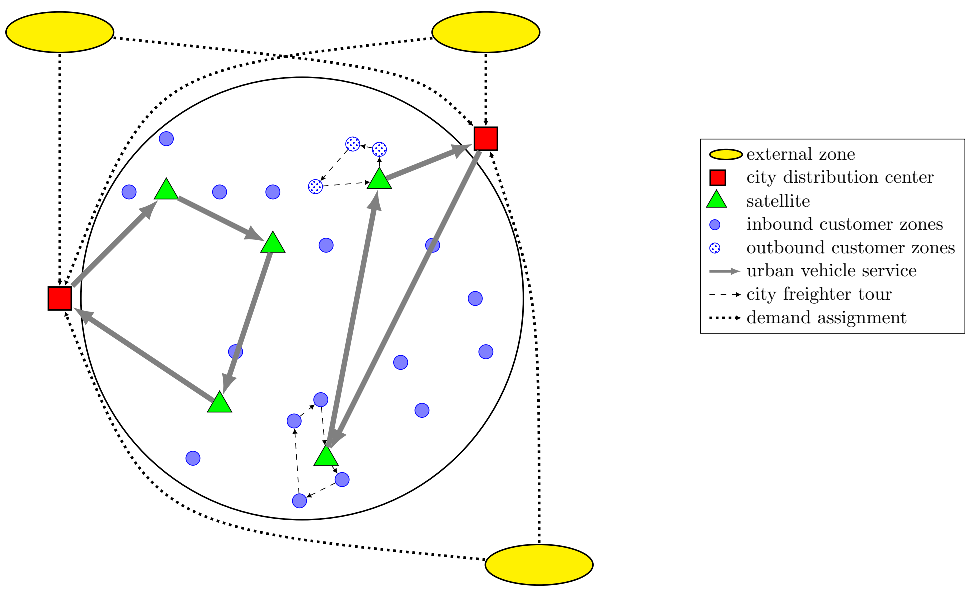

Figure 1: Two-tier City Logistics System

Inbound demand was the only demand type considered in most settings studied so far, that is, goods were assumed to move exclusively from locations outside the city, grouped for tactical-planning purposes into a small number of *external zones*, to customers in the city, the latter being grouped into *customer zones*. Crainic et al. (2012) examined the issue of inbound and outbound demands, and pointed out that their integration into a comprehensive plan promises important gains through better utilization of fewer vehicles, while also raising algorithmic challenges. The authors did not, however, propose any model nor any methodology to address these challenges. According to our best knowledge, we are the first to propose a model and solution methodology for the more general setting where inbound and outbound demands are jointly considered when planning the system activities.

Customer zones, represented by circular disks in Figure 1, and external zones, ellipse disks in the figure, may thus be both the origin and destination of demand. *Inbound demand* is to be delivered from external-zone origins to customers in the city. Symmetrically, *outbound demand* is to be picked up at customers in the city, to be shipped to specified external-zone destinations. Demand is characterized by a certain quantity of cargo to be moved from its origin to its destination. Demands often differ by commodity and might have restrictions on the vehicles used for the shipment (e.g., hazardous materials or products that need refrigeration). Crainic et al. (2009) showed how to model such situations. To simplify the presentation, and without loss of generality, we use the Physical Internet idea of multi-usage smart boxes that can be assembled for grouped shipping even when holding different products; such boxes are referred to as π-*containers* (Crainic and Montreuil, 2016). Demand is also characterized by time-related attributes and requirements, e.g., *availability times* and *time windows* associated to pickup and delivery activities, respectively, at facilities or customer locations.

The 2TM-CL supply network arrayed to answer demand is composed of facilities of two types, making up the the two tiers of the system, and the connections that transportation services of various modes provide among these facilities, the customer zones, and the external zones. *City distribution centers* (*CDC*s, also known as logistics platforms, urban distribution centers, and consolidation centers; illustrated through squares in Figure 1) make up the first tier. They are generally located at the outskirts of the city, close to major transportation axes and intermodal facilities. At CDCs, inbound and outbound cargo are sorted, consolidated, and prepared for distribution to customers in the city and long-haul transportation to final destinations at external zones, respectively.

Satellites, or satellite platforms (Crainic et al., 2004, 2009), make up the second tier of facilities. These intermediary facilities, illustrated through triangles in Figure 1, are generally located close to the city areas where traffic is tightly controlled and large vehicles are typically not allowed or not recommended to penetrate. City centers, historical neighborhoods, residential streets, and cultural or spiritual-dedicated spaces are a few examples of such city areas. Satellites are thus the meeting points of first- and second-tier vehicles, where cargo is transferred between *urban vehicles*, first-tier large vehicles offering economies of scale, and *city freighters*, second-tier vehicles adequate for the controlled zones mentioned above. Transfers are generally performed according to trans-dock principles, with no or little storage, between synchronized vehicles.

The general operating principle of 2TM-CL systems may then be described as follows.

Inbound demand goods are first delivered from external zones to a CDC by various transportation modes (as these modes are not within the planning and management scope of the City Logistics system, their modeling is not within the scope of the tactical planning considered in this paper). Inbound cargo is then sorted and consolidated into urban vehicles, each demand to be delivered to a particular satellite, where it will be unloaded and transferred, together possibly with cargo delivered by other vehicles, to city freighters which provide the last-mile delivery to customers. Symmetrically, outbound goods are picked up at customers by city freighters and brought to a specific satellite, where they are consolidated and transferred to an urban vehicle, which will bring them to the appointed CDC, to be then shipped on the long-haul journey to their external-zone destination. First and second-tier vehicles carrying inbound or outbound cargo could thus be present simultaneously at a satellite, competing for the capacity it offers for vehicle docking or parking, and for cargo transfer.

Different from most of the literature where trucking is the only transportation mode present, we consider a *multi/intermodal transportation system* (π-containers make the system intermodal). City freighters, operating at the second tier, may thus be eco-friendly vehicles such as electric small vans, traditional or cargo bikes, canal or river barges, etc. At the first-tier, the spectrum of possible modes is even larger as, increasingly, the utilization of what may be described as massive-flow modes and vehicles, e.g., regular and light-rail (Arvidsson and Browne, 2013), is contemplated or tested. We distinguish between *line-based* and *no-line* transportation modes and services. The latter include the various trucks and barges for which one may define services along any path within their admissible network (e.g., a city "trucking network" is often defined to restrict circulation). Line-based modes are often captive of particular infrastructure, such as passenger buses, which are "captive" of their predefined lines, regular rail, light rail (tramways, subways, etc.), and trolleybuses. This characteristic restricts the definition of line-based services to the network of the corresponding infrastructure. Thus, for example, a freight tramway service has to be defined on the network of tramway tracks, without being restricted to the tramway lines and stops operated for passengers (Trentini and Mahléné, 2010).

Two main approaches for line-based services are being contemplated within City Logistics projects around the world. On the one hand, regular vehicles may be equipped with special compartments for the transportation of goods (Lindholm and Behrends, 2012). On the other hand, freight-dedicated vehicles may be operated on the same infrastructure, either independently or as parts of regular convoys. The city of Paris, France, for example, is considering the first case, where freight tramways would be inserted in-between passenger tramways, but their loading and unloading would be performed at secondary stations, located on side lines, which would serve as satellites with short-term tramway parking and cargo holding and handling capacity (Freemark, 2011; Danard and Janin, 2016). A second example of such an intermodal transportation system is the Logistik-Tram in Frankfurt, Germany. In this pilot project, tramways deliver cargo to tramway stations (the satellites), where freight is unloaded and loaded into cargo bikes (Riemann, 2019; Brandt, 2018). Adding one or a few railcars to a regular train connecting a CDC sited along its line and downtown-located stations, which thus become satellites, illustrates the second case. Very little investigation of the integration of passenger and freight transportation has been performed so far, and then, only for particular services (Ghilas et al., 2013; Masson et al., 2017), but not in the general context of planning city-logistics services. This paper aims to start filling this gap.

Previous contributions generally assumed a fleet comprised of a single vehicle type at each tier, each vehicle having a single cargo space of a given capacity. While this is the case in a number of applications, it is certainly not the general situation. Multimodality comes with several different fleets at each tier. Moreover, even within the "same" mode, one may find different vehicle types in terms of capacity and cargo organization. With respect to this latter characteristic, several vehicle types have more than one cargo-holding space, as illustrated by the multiple cargo bays ("doors") of several proposed cargo tramways, and the (vertical or horizontal) separators that may be used within large trucks operating on the first tier.

We introduce the term *compartment* to refer to a particular vehicle-holding space (which is also part of the Logistiktram project, Riemann, 2019; Brandt, 2018). We then greatly extent upon current literature by assuming heterogeneous fleets of multi-compartment vehicles operating on the first tier of the 2TM-CL system. To simplify the presentation, but without loss of generality, we assume in the following that each fleet has a given *fleet size* limiting the number of vehicles of a given type operating out of each CDC, and that the *compartment capacity* is the same for all compartments of a vehicle of a given fleet (it morphs into the vehicle capacity for single-compartment vehicle types).

It is noteworthy that most previous work in the literature did not explicitly consider how to load vehicles. Indeed, given that the focus was on inbound demand only, vehicle loading was assumed appropriate for distribution, last-in-first-out policies being often assumed. When both inbound and outbound may be loaded into urban vehicles and city freighters, loading/unloading rules must be defined. We assume a *pseudo-backhaul* policy in this paper. As defined by Crainic et al. (2012), who were among the very first to study the integration of several types of demand into City Logistics planning, a pseudo-backhaul policy means that a vehicle completes the current type of activity before initiating a different one. This translates in the setting of this paper in a policy stating that one can start loading outbound demand in the compartment of a vehicle only once all inbound demand present in the compartment has been unloaded.

This policy is based on the idea that operations at satellites (customers as well) should be streamlined (Trentini and Mahléné, 2010). Indeed, capacity, time-window, and synchronization restrictions and requirements, as well as the goal of reducing the presence of vehicles in the city, implies efficient vehicle unloading and loading activities, which makes searching for and resorting of loads undesirable. Otherwise, long stopping times at satellites could cause, for example, congestion and delays in the entire network of a line-based mode.

Tactical planning aims to built the service network to support these activities: select first-tier services and their schedules, and determine demand assignment to facilities, services and compartments, to globally optimize the system performance. We discuss the tactical planning for 2TM-LC next.

2.2 Tactical Planning

Tactical planning for consolidation freight carriers aims to select and schedule services, together with the itineraries used to move freight flows from origins to destinations in the resulting service network. The goal is to satisfy the regular demand in the most cost- and resource-utilization efficient way possible, while satisfying the service-quality levels set by the carrier to answer customer requirements. The tactical plan is thus also generally yielding activity profiles of terminals and the resources required to support the selected services. The service network and plan is determined for a rather short period called *schedule length*, e.g., a day or a week, and it is then repeatedly applied over a certain *planning horizon*, the *season*, e.g., six months. Note that this decision process assumes that the major elements of the plan, the selected scheduled services (which are made available to the targeted group of customers) and the principal resource allocations to services and terminals, will not be modified during regular operations for the length of the planning horizon. Adjustments of the plan to actual demand is rather mostly performed through modifications to the routing of demand flows at operation time.

Scheduled service network design formulations defined over time-space graphs are generally used to model such problems(Crainic, 2000; Crainic and Kim, 2007). SSND belongs to the class of network design problems, where it is generally assumed that all potential arcs, services for the SSND case, are available. When applied to the first tier of City Logistics, this means the set of all potential line-based and no-line feasible services. Thus, for example, the cargo-tramway services that are feasible with respect to the passenger-tramway schedule are defined, as well as the feasible most efficient, in terms of the generalized transportation cost defined above, motor-carrier routes. The generally limited number of satellites simplifies the *a priori* computation of all potential services. If a very large number of satellites is contemplated, a dynamic service generation procedure (see, e.g., Boland et al., 2017; Hewitt, 2018) could be developed. Such algorithmic developments are beyond the scope of the present paper, however, and are left for future research.

In City Logistics terms, tactical planning targets the regular demand and is about selecting the services and resources that will be operated regularly and repeatedly. In this sense, first-tier service operations should be more stable, particularly for line-based services, as they move larger loads between CDCs and satellites. With this regularity comes the regularity in using the terminal facilities and the main customer-to-satellite assignment policy, while actual routing may vary from day to day according to the particular demand. With such a plan, material and human resources can be allocated for the duration of the planning horizon, which makes management easier and lowers costs, while still providing flexibility in day-to-day operations to instantiate the itineraries for the particular revealed demands.

The SSND formulation we detail in Section 4 is thus aimed at the main issues of tactical planning for 2TM-CL: 1) select a subset of scheduled services out of the set of possible line- and no-line-based multimodal services; 2) determine the itineraries of each inbound and outbound demand, including the assignment to a CDC, a satellite, and a particular service and, possibly, compartment; 3) manage the multimodal fleets and terminals. The goal is to determine the most cost-effective plan to satisfy forecasted demands with the available resources, where the generalized transportation costs account both for operations-related costs and for city-impact considerations (e.g., the relative non-suitability of certain streets or other public spaces for transport-related activities), as defined in Section 2.1.

3 Literature review

The coordination, planning and operation of multi-tier City Logistics transportation systems requires advanced planning methods. The literature is scarce, however, regarding both the system elements considered in the proposed models, as indicated above, and with respect to the solution methods aimed to address them efficiently, as shown in the detailed literature analysis of Bektaş et al. (2017) (see also Mancini et al., 2014; Taniguchi et al., 2014; Savelsbergh and Van Woensel, 2016).

Crainic et al. (2004) introduced a general two-tier city logistics system concept, together with a location-allocation methodology for the strategic decision issue of determining the satellite structure of the system. The authors considered inbound demand only and single motor-carrier modes and fleets for each of the two tiers of the system.

Crainic et al. (2009) proposed the first modeling framework for the short to medium-term planning of such single-mode, inbound-demand 2T-CL systems. The authors did consider, however, the possibility of different vehicle types with specific product-to-vehicle assignment rules. They also introduced to the City Logistics literature the time-dependency of demand and the corresponding issue of scheduling and synchronizing first-tier urban-vehicle services and second-tier city-freighter multi-tour working days. They proposed a path-based formulation for the day-before planning problem, when planning is performed shortly ("the day before") before operations when the demand is known.

The formulation combines scheduled service network design at the first tier and a time-dependent multi-tour synchronized vehicle routing problem with time windows at the second tier, by defining three sets of paths: the potential first-tier services, the potential working days for the second-tier vehicles of the heterogeneous fleet, and the demand itineraries from external zones to customers through satellites, which also enforced synchronization between first and second-tier fleets at satellites. The authors discussed solution-method avenues, introducing a meta-heuristic structure based on decomposing the problem along tiers, without actually solving the problem.

Crainic et al. (2016a) proposed a two-stage stochastic-programming formulation for a 2T-CL tactical planning model that explicitly accounts for demand uncertainty. The first stage corresponds to the selection of the first-tier services and the determination of the partial demand itineraries up to the selected satellites, as well as the satellite utilization in terms of customer assignments. The second stage models the selection of ad-hoc additional city freighters and the routing-based strategy to adjust the plan once demand is realized. The authors used the meta-heuristic of Crainic et al. (2009) as the basis of a Monte-Carlo evaluation procedure of several recourse strategies with increasing degrees of flexibility in routing and customer assignments.

Crainic et al. (2009) also made two observations. First, that one could view the service selection on the first tier as a particular case of a vehicle routing problem, yielding a two-tier, or, in VRP terminology, a two-echelon time-dependent, multi-tour, synchronized, vehicle routing problem with time windows, origin-destination (OD) demand, and heterogeneous fleets. Several publications followed focusing on two-echelon vehicle routing (e.g., Mancini et al., 2014; Perboli et al., 2011; Hemmelmayr et al., 2012; Contardo et al., 2012; Breunig et al., 2016) and location-routing (e.g., Boccia et al., 2010; Gianessi et al., 2016; Winkenbach et al., 2016; Boccia et al., 2017) problems. It is noteworthy that most of these contributions do not include the full range of attributes defined above; thus, OD demand, multiple demand types, time-dependency, multiple tours, and synchronization are generally missing. The second observation of Crainic et al. (2009) is that for longer-term planning, e.g., mid-term tactical planning and the evaluation of long-term strategic alternatives, the second-tier routing problem could be approximated and added to the first-tier formulation through appropriately-defined service costs on satellite-to-customer-zone links. Crainic and Sgalambro (2014) adopted this idea and focused on the modeling of the first-tier service network design within the day-before planning problem, studying the impact of a number of system parameters on the final design.

As this brief review illustrates, the contributions in the literature focused on simpler problem settings, none of the new features addressed in this paper having been previously included. The development of solution methods is also wanting. We present, in the following sections, a model and solution method that aim to start filling this gap.

4 General Approach and Model

As indicated in Section 2.2, our approach is to focus the tactical-plan model on the activities with long-term regularity, i.e., on the first tier, which includes the the first tier multimodal services, the assignments of demands (external zones) to CDCs, as well as the mode-specific assignments of demands (customer zones) to satellites. We approximate the corresponding routing costs, as indicated by Crainic et al. (2009).

The model selects a subset of scheduled services out of the set of potential line-based and no-line multimodal services, to satisfy the inbound and outbound demands at minimum total generalized transportation cost, which includes the costs of assigning to each demand a CDC and a satellite, and moving the corresponding freight between the respective facility and the corresponding external and customer zones. Services are characterized by their mode, route from a CDC through a number of satellites and back to the CDC, schedule, and a generalized transportation cost for each component of the route. Services may carry both inbound and outbound flows. They may have several compartments but only one type of flow is loaded in a given compartment at any time. The model takes into account several resources and fleet types, the available fleets at CDCs, and the capacities of satellites in terms of urban vehicles which can be unloaded/loaded simultaneously and the volume that can be transferred from/into city freighters at each time period.

We introduce the notation in Section 4.1, and present the MILP formulation in Section 4.2.

4.1 Notation and assumptions

We propose a SSND formulation on a time-space network. The schedule length is thus divided into periods $1, \ldots, T$. To somewhat simplify the presentation, we follow Crainic et al. (2009) and define the period length to be sufficiently small such that (1) at most one departure of a service from its city distribution center may take place during a period and, (2) all considered time-related parameters are integer multiples of the period length.

The set of city distribution centers $\mathcal{E}$ are the facilities that connect the external world and the city. This is where inbound goods are sorted and loaded into urban vehicles to be transported to satellites in set $\mathcal{Z}$. It is also where outbound goods delivered by urban vehicles from satellites are prepared and shipped to destinations outside the city. Notice that, in general, CDCs could also operate as satellites to close-by customers. First-tier services (i.e., urban vehicle services) are not needed for those customers and, thus, those customers are not part of the model.

Let $\mathcal{M}$ represent the set of transportation modes (trucks and tramways in the experiments of Section 6). Let $\mathcal{T}_m$ be the set of available urban-vehicle types for mode m (e.g., small trucks, large trucks, for the truck mode, and tramways with different numbers of compartments for the tramway mode), and $\mathcal{T} = \bigcup_m \mathcal{T}_m$ be the set of all urban-vehicle types. Then, for each urban-vehicle type $\tau \in \mathcal{T}$, let u^c_τ be its compartment capacity, and $n_{e\tau}$ be its fleet size at city distribution center $e \in \mathcal{E}$.

In most cases, city distribution centers are large facilities, where capacity issues are not critical and sufficient space is available for vehicles to wait for loading or unloading activities. This is, however, not the case for satellites, where the space available for transferring goods limits the number of urban vehicles and city freights which can be present simultaneously. Furthermore, there is generally no space available for storing goods at satellites, nor for vehicles to wait. The satellite capacity may also be time dependent, either due to opening hours given by the neighborhood, or to operations on a shared infrastructure. For example, a passenger tramway cannot wait longer than "normal" at a station because of unloading or loading activities of another tramway with freight somewhere down the line.

To account for such limitations, we introduce three different capacity measures for each satellite z at each period t: 1) a_{zt} for the total number of urban vehicles it may accommodate; 2) a^m_{zt} for the number of urban vehicles of mode $m \in \mathcal{M}$ it may accommodate; for trucks, this is the actual number of vehicles, while for tramways it is the number of available tracks (but could also be the number of cars or compartments); this definition allows multiple satellite use, e.g., tramway stops as tramway and truck satellites, as in the Paris concept (Danard and Janin, 2016); 3) b_{zt} for the total volume of goods the satellite may handle.

Demand is represented by two disjoint sets, $\mathcal{D}^I$ and $\mathcal{D}^O$, for inbound and outbound demand, respectively. When the same customer location is both the origin and the destination of demands, separate nodes are created, are assigned to the respective inbound and outbound demand sets, and are treaded individually within the model. Each demand $d \in \mathcal{D} = \mathcal{D}^I \cup \mathcal{D}^O$ is defined by its origin and destination, volume v_d, and time windows indicating when it is available at origin and when it must be delivered at destination. In the network model we propose, these time windows are defined with respect to the nearest CDC (for external zones) and satellite (for customer zones). As the model also selects the CDC and satellite for each demand, the time windows are adjusted for each potential facility, e.g., when an inbound demand is assigned to a different CDC from its closest one, the origin time window for the selected CDC is shifted in time to account for the additional transport activities.

External zones, the out-of-city origins and destinations, are linked by various transportation modes to the city and the CDCs. Projecting these locations onto the nearest (by some travel measure) CDC appears not necessarily to be the best global decision.

This intuition has been validated by Gianessi (2014) and Gianessi et al. (2016). Therefore, we define for each demand $d \in \mathcal{D}$ a set of potential CDCs $\mathcal{E}(d) \subseteq \mathcal{E}$, and associate a cost f_{de} for assigning demand d to CDC $e \in \mathcal{E}(d)$ to account for using another CDC rather than the closest one or for using inter-CDC transportation. Similarly, we let the model decide on which satellite to use for each demand. To streamline the network representation and the model, however, we do not add satellite-customer arcs; we rather add the corresponding cost to the cost of the service carrying the flow into or out of the satellite. Let $\mathcal{Z}(d) \subseteq \mathcal{Z}$ be the set of satellites that may service demand d. We then define, for each satellite $z \in \mathcal{Z}(d)$ and service r (that could service the demand in time), the satellite-customer transportation cost s_{dzr} as the approximated cost of delivery to or pick up from the customer of demand d. The assignment costs s_{dzr} represent not only the transport, unloading, and loading costs, but also city-disturbance factors related to these activities.

Transportation in the first tier is performed by urban-vehicle *services* to be selected from a given set $\mathcal{R}$. Service $r \in \mathcal{R}$ is characterized by mode m_r, vehicle type τ_r, cost k_r, origin, destination, sequence of satellites visited, and their schedule. The cost k_r represents not only the operating costs of circulating, unloading,and loading, but also city-disturbance factors related to these activities.

A given service $r \in \mathcal{R}$ starts at CDC $e_r \in \mathcal{E}$, visits several satellites $\mathcal{Z}(r)$, and returns to the same city distribution center. Each service has a defined time schedule that considers the travel times, the service times, and possible waiting times. $\mathcal{R}(z, t)$ ($\mathcal{R}(z, t, m)$) defines the subset of services which are present at and using the resources of satellite z at time t (of mode m), while $\mathcal{R}(t, \tau, e)$ is the subset of services of vehicle-type τ, operated out of CDC e at period t.

Vehicles have compartments. Therefore, for each service $r \in \mathcal{R}$, $\mathcal{R}^C(r)$ defines the set of *compartment services*. Each element of the set reflects a given vehicle compartment of service r. For single-compartment services, $|\mathcal{R}^C(r)| = 1$. Also, when the service $r \in \mathcal{R}$ is operated, all its corresponding compartment services are operated as well. For operational convenience, the model prohibits the simultaneous assignment of inbound and outbound flows to the same compartment of a service. Therefore, we assume that a compartment is first used to offload inbound demand and then could be used to upload outbound demand.

We conclude this part by pointing out that the combined definition of potential services $\mathcal{R}$ and satellite capacities can be used to restrict access of particular vehicle types to particular facilities or during given time periods. Thus, for example, when a tramway track is used exclusively by public transportation, or a satellite (or its neighborhood) is not available during specific time intervals, the case is easily addressed by not defining services using that infrastructure at that time. A summary of the notation is given in Table 7 in the Appendix.

4.2 Formulation

Crainic et al. (2009) formulated a three-path SSND model, the set of demand itineraries (paths) ensuring the synchronization between the the first-tier and second-tier activities and their respective path representations. We introduce a new formulation where paths are no longer necessary for synchronization purposes. The formulation has many similarities with knapsack and bin packing problems and, thus, good bounds and efficient solution methods can be devised.

An itinerary of an inbound demand is illustrated in Figure 2. The figure shows an urban-vehicle service starting from a CDC (square in the figure) and visiting three satellites (triangles) before returning to the CDC. The itinerary starts at the external zone (ellipse) from where the goods are received at the selected CDC (the dotted arc into the square). The considered inbound goods are loaded at the CDC into a compartment of the urban-vehicle service, and transported to the selected satellite, the second in the figure, from where they are to be delivered to the final customer zone (the dotted arc to the small disk). The itinerary is thus the sequence consisting of the CDC, the visited satellites of the service up to the second one, and the customer zone (including all movement arcs and relevant time stamps). One observes that, in the current setting, all the required information can be found through the selection of the service, compartment, and satellite. Selecting the service implies selecting the CDC for the delivery from the external zone, and the departure time. One thus does not need an explicit decision variable for the selection of the CDC. Similarly, selecting the satellite, which is on the route of the selected service, determines when the demand is ready to be delivered and how it will be delivered to the customer zone, while taking care of the synchronization issue. Then, as each demand itinerary is completely defined by assigning it to a service, a compartment, and a satellite, one can work simply with arc decision variables corresponding to this assignment. A similar discussion can be made for outbound demands.

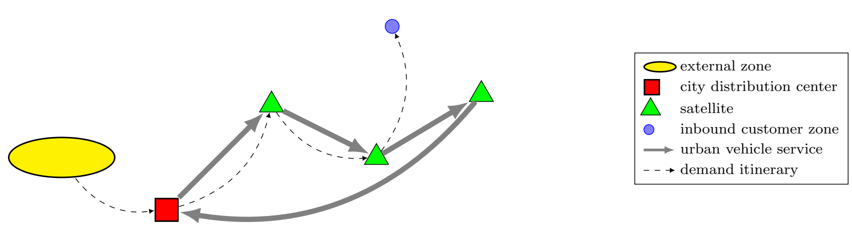

Figure 2: Illustration of itinerary vs. assignment decision

We therefore define the following two sets of binary decision variables:

- ρ_r taking value one if the urban-vehicle service $r \in \mathcal{R}$ is selected, and zero otherwise;
- $x_{r^c,d,z}$ taking value one if demand $d \in \mathcal{D}$ is assigned to compartment service $r_c \in \mathcal{R}^C$

and satellite $z \in \mathcal{Z}(r)$, and zero otherwise.

We then formulate the SSND problem as

$$\min \sum_{r\in\mathcal{R}} k_r \rho_r + \sum_{d\in\mathcal{D}} \sum_{r\in\mathcal{R}} \sum_{r^c\in\mathcal{R}^C(r)} \sum_{z\in\mathcal{Z}(r)} (s_{dzr} + f_{d,e_r})\, x_{r^c,d,z} \tag{1}$$

subject to

$$\sum_{r\in\mathcal{R}} \sum_{r^c\in\mathcal{R}^C(r)} \sum_{z\in\mathcal{Z}(r)} x_{r^c,d,z} = 1 \qquad \forall d \in \mathcal{D} \tag{2}$$

$$x_{r^c,d_1,z_1} + x_{r^c,d_2,z_2} \le 1 \qquad \forall r^c \in \mathcal{R}^C(r), r \in \mathcal{R}, d_1 \in \mathcal{D}^I, d_2 \in \mathcal{D}^O, z_1, z_2 \in \mathcal{Z}(r), z_1 \ge z_2 \tag{3}$$

$$\sum_{d\in\mathcal{D}^I} \sum_{z\in\mathcal{Z}(r)} v_d x_{r^c,d,z} \le u^c_{\tau_r} \rho_r \qquad \forall r^c \in \mathcal{R}^C(r), r \in \mathcal{R} \tag{4}$$

$$\sum_{d\in\mathcal{D}^O} \sum_{z\in\mathcal{Z}(r)} v_d x_{r^c,d,z} \le u^c_{\tau_r} \rho_r \qquad \forall r^c \in \mathcal{R}^C(r), r \in \mathcal{R} \tag{5}$$

$$\sum_{r\in\mathcal{R}(t,\tau,e)} \rho_r \le n_{e\tau} \qquad \forall \tau \in \mathcal{T}, e \in \mathcal{E}, t = 1, \ldots, T \tag{6}$$

$$\sum_{r\in\mathcal{R}(z,t)} \rho_r \le a_{zt} \qquad \forall z \in \mathcal{Z}, t = 1, \ldots, T \tag{7}$$

$$\sum_{r\in\mathcal{R}(z,t,m)} \rho_r \le a^m_{zt} \qquad \forall z \in \mathcal{Z}, m \in \mathcal{M}, t = 1, \ldots, T \tag{8}$$

$$\sum_{r\in\mathcal{R}(z,t)} \sum_{r^c\in\mathcal{R}^C(r)} \sum_{d\in\mathcal{D}} v_d x_{r^c,d,z} \le b_{zt} \qquad \forall z \in \mathcal{Z}, t = 1, \ldots, T \tag{9}$$

$$\rho_r \in \{0,1\} \qquad \forall r \in \mathcal{R} \tag{10}$$

$$x_{r^c,d,z} \in \{0,1\} \qquad \forall d \in \mathcal{D}, r^c \in \mathcal{R}^C(r), r \in \mathcal{R}, z \in \mathcal{Z}(r) \tag{11}$$

The objective function (1) minimizes the total generalized cost of selecting and operating services that move inbound and outbound demand flows, distributing demands from satellites and bringing outbound demands to satellites, as well as selecting a CDC for each demand. Constraints (2) ensure that each item is assigned exactly to one compartment, while Constraints (3) ensure that outbound demand is only assigned to a compartment after the inbound demand is unloaded and the compartment is empty. The compartment capacities for inbound and outbound traffic are enforced by the linking Constraints (4) and (5). These constraints combined with Constraints (3) enforce the capacity restriction for the entire service. Then, Constraints (6) ensure that the maximum number of vehicles of each type assigned to a city distribution center is never exceeded. Note that in these constraints, the first summation accounts for the service remaining at the satellite during the duration of its handling time. Constraints (7) and (8) limit the number of urban vehicles present at a satellite at each period in total and per transportation mode, respectively. Finally, Constraints (9) limit the amount of demand that can be unloaded or loaded at a satellite at each period.

This network design formulation presents a number of similarities to bin packing and knapsack problems. First, the loading of inbound demand and unloading of outbound demand at CDCs are close to two bin packing problems (Dyckhoff and Finke, 1992), where the compartments of the services are bins to which the demands, as items, are assigned to. Without Constraint (3), the bin capacity constraints are ensured throughout the whole service if they are satisfied at the city distribution center. This also implies that, for a known service schedule, the flow-optimization subproblem displays similarities with a multiple knapsack problem with assignment restrictions (Dawande et al., 2000), the latter accounting for the inbound and outbound conflicts and ensuring demand satisfaction. The main difference between our model and the knapsack and bin packing problems is the complex cost structure. Each service (bin) has a unique operating cost and each assignment of an item to a service also has a unique cost. Yet, we take advantage of the aforementioned similarities in the solution method we propose, which is presented next.

5 Solution Method

We use Benders decomposition (Benders, 1962) to propose an efficient exact solution method for the model proposed in the previous section. Although Benders decomposition is a solution method for general linear programs, several authors adapted it for integer programs (e.g., Laporte and Louveaux, 1993; Laporte et al., 2002; Restrepo et al., 2015). An overview of the method and associated acceleration techniques is given by Rahmaniani et al. (2017).

The fundamental idea in Benders decomposition is to decompose the problem into two easier-to-address subproblems. This is obtained by projecting the original problem onto the space defined by a subset of variables that are considered as complicating. The problem is then reformulated via the application of an outer linearization of the projected term and the resulting reformulation is solved via a relaxation method. Thus, a *master problem*, that includes both the complicating variables and a lower bounding value of the projected term, is successively solved to generate a lower bound for the original problem. The master problem is also updated iteratively via the inclusion of cuts, i.e., optimality cuts that express the value of the projected term and feasibility cuts that eliminate infeasible solutions with respect to the complicating variables. These cuts are generated through the solution of the *slave problem*, whose feasible solutions also enable to obtain an upper bound on the original problem.

Although Benders decomposition defines a general strategy that can be applied to a wide variety of optimization problems, as detailed in the literature review of Rahmaniani et al. (2017), the method requires enhancements to be implemented in order to run efficiently. In particular, the relaxation method that is prescribed when applying Benders decomposition may suffer from both ineffective initial iterations and instability throughout the solution process Rahmaniani et al. (2017). These issues stem from the initial relaxation that is applied to the projected term, which eliminates from the master problem all information regarding the slave problem. Considering that this information is only gradually added through the cut generation process, a large number of back and forth exchanges (iterations) may be required between the master and slave problems before the overall search converges. To alleviate these issues, we show that the decomposition we propose enables to obtain tight feasibility and optimality cuts from the slave problem. Furthermore, we propose two specialized enhancements to the decomposition strategy. First, we develop a set of valid inequalities and a novel partial decomposition technique inspired by Crainic et al. (2016b) that considerably strengthen the quality of the relaxation defined by the master problem. The latter enhancement actually extends the original partial decomposition strategy proposed for stochastic models (Crainic et al., 2016b), by applying general aggregation techniques (Tsurkov, 2013) to improve the quality of the lower bound that is provided by the master's formulation in the Benders method.

The remainder of this section is divided as follows: we first introduce the general solution procedure and explain how we handle the integer subproblem (Section 5.1). We then define the slave problem (Section 5.2), followed by the master problem (Section 5.3). In Section 5.3, we also detail the cuts generated from the slave problem and clearly present the specialized enhancements proposed.

5.1 Benders Decomposition for Integer Problems

The complicating variables in our problem are the ρ_r, which select services and establish the schedule. The easier variables are the demand assignment variables $x_{r^c,d,z}$. The master problem therefore selects the services to generate a lower bound ξ_{low}, while the slave problem solves a multiple knapsack problem with precedence constraints, which are due to the inbound-outbound loading restrictions. Our method benefits from the fact that the linear relaxation of knapsack problems gives very good lower bounds. Therefore, we can generate tight feasibility and optimality cuts through the linear relaxation of the slave problem. Moreover, we derive tighter lower bounds for the original integer problem when solving the master problem.

In the slave problem, we first solve the dual subproblem DSP for a given selection of services $\bar{\rho}$ to derive the classical Benders cuts. This solution gives an upper bound ξ_{DSP} to the relaxed problem. When this upper bound is worse than the current best integer upper bound ξ_{up}, the integer problem will also be worse and, therefore, the integer node is removed. We solve the integer subproblem SP when the relaxed upper bound is better, which yields a potential new upper bound ξ_{SP}. Because of the good linear relaxation of the integer slave problem, we can avoid solving too many integer subproblems. When the relaxed upper bound is better, the possibility of also getting a better true integer upper bound is high. To improve the convergence of the relaxed problem, we further add the combinatorial cuts and additional valid inequalities described in Section 5.3.

Algorithm 1 Solution procedure

Set $\xi_{low} \leftarrow -\infty, \xi_{up} \leftarrow \infty$
while $\xi_{up} > \xi_{low}$ **do**
    solve $RMP \rightarrow \bar{\rho}, \xi_{low}$
    update core point (Papadakos, 2008)
    solve $DSP(\bar{\rho}) \rightarrow \xi_{DSP}$
    if DSP is bounded **then**
        generate optimality cut (Magnanti and Wong, 1981)
        if $\xi_{DSP} < \xi_{up}$ **then**
            solve $SP(\bar{\rho}) \rightarrow \xi_{SP}$
            if $\xi_{SP} < \xi_{up}$ **then**
                $\xi_{up} \leftarrow \xi_{SP}$
            end if
        end if
    else
        generate feasibility cut
    end if
    generate combinatorial cut and additional valid inequalities
end while

Algorithm 1 summarizes the general structure of the solution procedure based on Benders decomposition. To further accelerate the solution method, the used procedure is embedded into a Branch-and-Cut framework. At each integer node with a potential better solution, the subproblem is solved and cuts are generated.

5.2 Slave Problem

The slave problem $SP(\bar{\rho})$ for a given selection of services $\bar{\rho}$ is defined as follows:

$$\min \sum_{d\in\mathcal{D}}\sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)}\sum_{z\in\mathcal{Z}(r)} (s_{dzr} + f_{d,e_r})\, x_{r^c,d,z} \tag{12}$$

subject to

$$\sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)}\sum_{z\in\mathcal{Z}(r)} x_{r^c,d,z} = 1 \qquad \forall d\in\mathcal{D} \tag{2}$$

$$x_{r^c,d_1,z_1} + x_{r^c,d_2,z_2} \le 1 \qquad \forall r^c\in\mathcal{R}^C(r), r\in\mathcal{R}, d_1\in\mathcal{D}^I, d_2\in\mathcal{D}^O, z_1,z_2\in\mathcal{Z}(r), z_1\ge z_2 \tag{3}$$

$$\sum_{r\in\mathcal{R}(t,z)}\sum_{r^c\in\mathcal{R}^C(r)}\sum_{d\in\mathcal{D}} v_d x_{r^c,d,z} \le b_{zt} \qquad \forall z\in\mathcal{Z}, t=1,\dots,T \tag{9}$$

$$\sum_{d\in\mathcal{D}^I} v_d x_{r^c,d,z} \sum_{z\in\mathcal{Z}(r)} \le u^c_{\tau_r}\bar{\rho}_r \qquad \forall r^c\in\mathcal{R}^C(r), r\in\mathcal{R} \tag{13}$$

$$\sum_{d\in\mathcal{D}^O} v_d x_{r^c,d,z} \sum_{z\in\mathcal{Z}(r)} \le u^c_{\tau_r}\bar{\rho}_r \qquad \forall r^c\in\mathcal{R}^C(r), r\in\mathcal{R} \tag{14}$$

$$x_{r^c,d,z}\in\{0,1\} \qquad \forall d\in\mathcal{D}, r^c\in\mathcal{R}^C(r), r\in\mathcal{R}, z\in\mathcal{Z}(r) \tag{15}$$

The dual problem of the linear relaxation of $SP(\bar{\rho})$, $DSP(\bar{\rho})$, is solved to generate optimality and feasibility cuts. Let $\alpha(d)$, $\beta(d_1,d_2,r^c,r,z_1,z_2)$, $\delta(z,t)$, $\gamma^{In}(r,r^c)$ and $\gamma^{Out}(r,r^c)$ be the dual variables of constraints (2),(3),(9), (13) and (14), respectively. The dual slave problem is then defined as follows:

$$\max \sum_{d\in\mathcal{D}}\alpha(d) + \sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)}\sum_{d_1\in\mathcal{D}^I}\sum_{d_2\in\mathcal{D}^O}\sum_{z_1\in\mathcal{Z}(r)}\sum_{\substack{z_2\in\mathcal{Z}(r)\\ z_1\ge z_2}}\beta(d_1,d_2,r^c,r,z_1,z_2)$$
$$+\sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)} u^c_{\tau_r}\bar{\rho}_r\left(\gamma^{In}(r,r^c)+\gamma^{Out}(r,r^c)\right) + \sum_{z\in\mathcal{Z}}\sum_{t=1}^{T} b_{zt}\delta(z,t) \tag{16}$$

subject to

$$\alpha(d) + \sum_{d_2 \in \mathcal{D}^O} \sum_{\substack{z_2 \in \mathcal{Z}(r) \\ z \geq z_2}} \beta(d, d_2, r^c, r, z, z_2) + v_d \left(\gamma^{In}(r, r^c) + \gamma^{Out}(r, r^c)\right)$$
$$+ \sum_{\substack{t=1| \\ r \in \mathcal{R}(t,z)}}^{T} v_d \delta(z, t) \leq s_{dzr} + f_{d,e_r}$$
$$\forall d \in \mathcal{D}^I, r^c \in \mathcal{R}^C(r), r \in \mathcal{R}, z \in \mathcal{Z}(r) \quad (17)$$

$$\alpha(d) + \sum_{d_1 \in \mathcal{D}^I} \sum_{\substack{z_1 \in \mathcal{Z}(r) \\ z_1 \geq z}} \beta(d_1, d, r^c, r, z_1, z) + v_d \left(\gamma^{In}(r, r^c) + \gamma^{Out}(r, r^c)\right)$$
$$+ \sum_{\substack{t=1| \\ r \in \mathcal{R}(t,z)}}^{T} v_d \delta(z, t) \leq s_{dzr} + f_{d,e_r}$$
$$\forall d \in \mathcal{D}^O, r^c \in \mathcal{R}^C(r), r \in \mathcal{R}, z \in \mathcal{Z}(r) \quad (18)$$

$$\alpha \in \mathbb{R} \quad (19)$$

$$\beta, \gamma^{In}, \gamma^{Out}, \delta \leq 0 \quad (20)$$

Since the primal slave problem consists of one set of decision variables ($x_{r^c,d,z}$), the dual problem has only one set of constraints. Because of the precedence constraints (3), we separate the constraints into one set for the inbound demand and one for the outbound demand. Assuming that $\bar{\rho}$ defines a feasible selection of services, then the cut obtained using the solution of (16) - (20) can be further improved. Specifically, a Pareto optimal cut can be derived by resolving the slave problem, where the objective is redefined using a core-point of the master problem's feasible region, and where a constraint is added to ensure that the obtained solution is equal in value to the previous one (see Magnanti and Wong, 1981). The necessary core-points are updated according to Papadakos (2008).

5.3 Master Problem

Using the information provided by the slave problem, we can derive at each integer node either an optimality cut, if the dual slave problem is bounded, or a feasibility cut, otherwise.

Let α^*, β^*, δ^*, γ^{In*} and γ^{Out*} be the optimal solutions of the dual slave problem, and

ξ a continuous decision variable. The resulting optimality cuts are:

$$\sum_{d\in\mathcal{D}}\alpha^*(d)+\sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)}\sum_{d_1\in\mathcal{D}^I}\sum_{d_2\in\mathcal{D}^O}\sum_{z_1\in\mathcal{Z}(r)}\sum_{\substack{z_2\in\mathcal{Z}(r)\\z_1\geq z_2}}\beta^*(d_1,d_2,r^c,r,z_1,z_2)$$
$$+\sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)}u^c_{\tau_r}\rho_r\left(\gamma^{In*}(r,r^c)+\gamma^{Out*}(r,r^c)\right)+\sum_{z\in\mathcal{Z}}\sum_{t=1}^{T}b_{zt}\delta^*(z,t)\leq\xi \tag{21}$$

In case of an unbounded dual slave problem, the extreme rays α^*, β^*, δ^*, γ^{In*} and γ^{Out*} give the following feasibility cut

$$\sum_{d\in\mathcal{D}}\alpha^*(d)+\sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)}\sum_{d_1\in\mathcal{D}^I}\sum_{d_2\in\mathcal{D}^O}\sum_{z_1\in\mathcal{Z}(r)}\sum_{\substack{z_2\in\mathcal{Z}(r)\\z_1\geq z_2}}\beta^*(d_1,d_2,r^c,r,z_1,z_2)$$
$$+\sum_{r\in\mathcal{R}}\sum_{r^c\in\mathcal{R}^C(r)}u^c_{\tau_r}\rho_r\left(\gamma^{In*}(r,r^c)+\gamma^{Out*}(r,r^c)\right)+\sum_{z\in\mathcal{Z}}\sum_{t=1}^{T}b_{zt}\delta^*(z,t)\leq 0 \tag{22}$$

Let $C^O=\{c_o\}$ and $C^F=\{c_f\}$ be the current sets of optimality and feasibility cuts, respectively. We also generate a combinatorial cut

$$\bar{\xi}_{\bar{\rho}}\left(\sum_{r\in\mathcal{R}|\bar{\rho}_r=1}\rho_r-\sum_{r\in\mathcal{R}|\bar{\rho}_r=0}\rho_r-\sum_{r\in\mathcal{R}}\bar{\rho}_r+1\right)\leq\xi \tag{23}$$

at each integer node $\bar{\rho}$ with costs $\bar{\xi}_{\bar{\rho}}$ of the subproblem. This cut is added to the set of combinatorial cuts $C^C=\{c_c\}$. Therefore, for a given feasible selection of services, this combinatorial cut will bound the value of ξ to the associated total assignment costs of the demands to the satellites through time and to the city distribution centers (i.e., the corresponding optimal value of the slave problem).

5.3.1 Valid Inequalities

To further strengthen the master problem relaxation, we introduce four sets of inequalities: two for inbound demands, inequalities (24) and (25), and two for outbound demands, inequalities (26) and (27):

$$\sum_{d \in \mathcal{D}^{I-}(t)} v_d - \sum_{r \in \mathcal{R}^{I-}(t)} u_{\tau_r} \rho_r \leq 0, \qquad t = 1, \ldots, T \tag{24}$$

$$\sum_{d \in \mathcal{D}^{I+}(t)} v_d - \sum_{r \in \mathcal{R}^{I+}(t)} u_{\tau_r} \rho_r \leq 0, \qquad t = 1, \ldots, T \tag{25}$$

$$\sum_{d \in \mathcal{D}^{O-}(t)} v_d - \sum_{r \in \mathcal{R}^{O-}(t)} u_{\tau_r} \rho_r \leq 0, \qquad t = 1, \ldots, T \tag{26}$$

$$\sum_{d \in \mathcal{D}^{O+}(t)} v_d - \sum_{r \in \mathcal{R}^{O+}(t)} u_{\tau_r} \rho_r \leq 0, \qquad t = 1, \ldots, T \tag{27}$$

$\mathcal{D}^{I-}(t) \subseteq \mathcal{D}$ defines the subset of inbound demands arriving at their destination in period t or earlier ($\mathcal{D}^{I-}(t_1) \subseteq \mathcal{D}^{I-}(t_2)$ for $t_1 \leq t_2$). Since the arrival period of a demand may differ depending on the satellite used for the final distribution, the latest out of these period is used. Similarly, $\mathcal{R}^{I-}(t) \subseteq \mathcal{R}$ are the services which can operate these inbound demands with period t. The volume of all demands until period t must then be lower than the capacity of all operated services which can satisfy these demands. This is ensured by valid inequality (24). Instead of inbound dates, inequality (25) considers the availability at city distribution centers. $\mathcal{D}^{I+}(t) \subseteq \mathcal{D}$ are the inbound demands that are available in period t or later. Therefore, $\mathcal{D}^{I+}(t_1) \supseteq \mathcal{D}^{I+}(t_2)$ for $t_1 \leq t_2$ holds. This ensures that sufficient capacity is also available for resources which are available in the last period only. Valid inequalities (26) and (27) ensure the same relations for the outbound demands, where the arrival periods at CDCs and the availability at satellites are considered.

5.3.2 Partial Benders Decomposition

Inspired by the idea of partial decomposition for stochastic programming problems (Crainic et al., 2016b), we propose to retain some information regarding the subproblem in the master problem. However, instead of adding a particular set of scenario subproblems as proposed by Crainic et al. (2016b), we retain a relaxed version of the slave problem that is obtained by applying a row aggregation (Tsurkov, 2013) over the compartments, while taking advantage of the fact that each demand is assigned to exactly one compartment of one service only. The aggregated decision variable $\hat{x}_{rdz}$ reflects the sum over all compartments $r^c \in \mathcal{R}(r)$ of $x_{r^c,d,z}$ for all demands $d \in \mathcal{D}$, satellites $z \in \mathcal{Z}$, and services $r \in \mathcal{R}$. For a given feasible selection of services, this aggregation thus produces a relaxed slave problem where the number of integer variables is reduced while still providing a great deal of the information contained in the original slave model. The relaxed master problem can then be strengthened by reformulating it as a mixed-integer linear program (as opposed to a pure integer program). This is achieved by modifying the decomposition approach that is used to solve the original problem (1) - (11). Specifically, the linear relaxation of the aggregated slave problem is first added to the model (1) - (11) as a set of redundant variables and constraints. The Benders decomposition strategy can then be applied as stated before, the only addition in this case being that the set of redundant continuous variables are included in the projection (i.e., add the redundant continuous variables to the set of complicating variables). Then, by performing the subsequent steps of Benders decomposition, we obtain a new current relaxed master problem, as defined by (28) - (39). Similarly to the stochastic programming case (Crainic et al., 2016b), this additional information helps guiding the algorithm towards good master solutions.

We thus redefine the current relaxed master problem RMP as:

$$\min \sum_{r \in \mathcal{R}} k_r \rho_r + \xi \tag{28}$$

subject to

$$\text{Constraints } (6), (7), \text{ and } (8)$$

$$\text{Valid inequalities } (24), (25), (26), \text{ and } (27)$$

$$c_o \qquad \forall c_o \in C^O \tag{29}$$

$$c_f \qquad \forall c_f \in C^F \tag{30}$$

$$c_c \qquad \forall c_c \in C^C \tag{31}$$

$$\sum_{d \in \mathcal{D}} \sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{Z}(r)} (s_{dzr} + f_{d,e_r}) \, \hat{x}_{rdz} \leq \xi \tag{32}$$

$$\sum_{r \in \mathcal{R}} \sum_{z \in \mathcal{Z}(r)} \hat{x}_{rdz} = 1 \qquad \forall d \in \mathcal{D} \tag{33}$$

$$\sum_{d \in \mathcal{D}^I} \sum_{z \in \mathcal{Z}(r)} v_d \hat{x}_{rdz} \leq n^c_{\tau_r} u^c_{\tau_r} \rho_r \qquad \forall r \in \mathcal{R} \tag{34}$$

$$\sum_{d \in \mathcal{D}^O} \sum_{z \in \mathcal{Z}(r)} v_d \hat{x}_{rdz} \leq n^c_{\tau_r} u^c_{\tau_r} \rho_r \qquad \forall r \in \mathcal{R} \tag{35}$$

$$\sum_{r \in \mathcal{R}(t,z)} \sum_{d \in \mathcal{D}} v_d \hat{x}_{rdz} \leq b_{zt} \qquad \forall t = 1, \ldots, T, z \in \mathcal{Z} \tag{36}$$

$$0 \leq \hat{x}_{rdz} \leq 1 \qquad \forall d \in \mathcal{D}, r \in \mathcal{R}, z \in \mathcal{Z}(r) \tag{37}$$

$$\xi \in \mathbb{R} \tag{38}$$

$$\rho_r \in \{0, 1\} \qquad \forall r \in \mathcal{R} \tag{39}$$

We minimize the service operating costs subject to the service related constraints from the IP formulation and the generated cuts. Constraints (32) - (37) are related to the partial decomposition. The costs of the partial decomposition give a lower bound to the objective function through Constraint (32). Constraints (34) and (35) enforce the capacity restrictions, where the compartment capacity is multiplied with the number of compartments to reflect the total capacity of the service. Constraints (33) and (36) reflect the demand assignment and the satellite capacity restrictions as before.

Through the inclusion of Constraints (33) - (37), this strategy has the added advantage of providing explicit requirements that need to be enforced on the selected services with respect to how the demands can be feasibly assigned to them (i.e., requirements that in the original Benders decomposition were only enforced in the slave problem). Therefore, we expect that this will mitigate the feasibility issues that may be experienced at the master problem level (i.e., identify feasible solutions to the master problem more rapidly) and thus improve the upper bound generated by the solution method. Furthermore, the inclusion of the inequality (32) defines another lower bound on the value of ξ, which complements the bound that is provided through both the optimality and combinatorial cuts contained in the model. By simultaneously improving both the upper and lower bounds, we expect that this strategy will greatly increase the overall efficiency of the solution process.

6 Numerical Study

We performed a series of numerical experiments to, first, evaluate the performance of the proposed solution method and, second, evaluate the impact of including multimodality and multiple demand types into tactical planning models on system performance measures.

We first describe the numerical setup in Section 6.1. We then examine the performance of the proposed solution method, in Section 6.2, through a comparison with a commercial solver. Insights into benefits of using multimodal transportation systems are discussed in Section 6.3, while the importance of considering inbound and outbound demands in the same network model is shown in Section 6.4.

6.1 Numerical Setup

All experiments were conducted on a cluster of 27 machines each having two Intel(R) Xeon(R) X5675 3.07 GHz processors with 96 GB of RAM running on Linux. Each machine has 12 cores and each experiment was run using a single thread. Both the IP formulation and the proposed Benders Decomposition algorithm (BD) were implemented in C++ using CPLEX Concert Technology 12.6.1. CPLEX 12.6.1 is used as the benchmark solver.

We generated a set of new instances based on four networks illustrated in Appendix 8.2. The networks are inspired by a typical city structure and by the work of Crainic and Sgalambro (2014), but the data is randomly generated. Each network has two CDCs. One has four satellites, one six, and two have eight. We used the Euclidian distances between all points and assumed different travel speeds depending on the vehicle used. We considered a planning horizon of 3 hours divided into 36 periods of 5 minutes each. Each satellite has a capacity of 5,000 (u^{V}_{zt}) and can process one truck or one tramway per period ($u^{\mathcal{T}}_{zt} = 1$), thus implying that $u^{m}_{zt} \leq 1$ for all modes.

We generated tramway services on predefined lines (shown in the network illustrations) and truck services. We generated two tramway lines, one for each city distribution center, for networks 1–3, and one line for network 4. The tramway has three compartments with a capacity of 700 each (u^{c}_{τ}). It visits a set of satellites and then returns on the same track in the opposite direction. The tramway is operated several times during the schedule length, each departure defining a service. We assume that the tramway moves on dedicated infrastructure and does not have to deal with congestion. Therefore, its travel speed is higher than for the truck. For the truck services, a subset of satellites and a city distribution center is selected for each, while the starting period of the service is generated randomly. For a basic scenario, we assumed a large truck with a capacity of 3,000 (u_{τ}). This resulted in 73, 71, 69, and 70 services for the four networks, respectively. When adding additional services for larger instances to the network, a smaller truck with a capacity of 2,000 (u_{τ}) is further added. Finally, a fixed selection cost, as well as a total distance-depending generalized transportation cost was calculated and assigned to each service.

We generated 24 demand instances with 150, 160, 170, and 180 demands and different shares of inbound and outbound demands: 50%, 67% and 83% inbound demand out of the total demand. We generated two instances for each combination of demand and inbound and outbound share. These instances were associated to each of the four networks leading to a total of 96 instances. Demand zones are randomly distributed ("over the city center"). Each demand is either an inbound or an outbound demand and has assignment costs for each city distribution center. A volume between 50 and 100 was randomly generated, as well as time windows. Considering inbound demands, we generated delivery time windows at satellites and pickup time windows for the availability at the CDCs. Considering outbound demands, we generated pickup time windows at satellites and delivery time windows for the CDCs. We defined the customer-satellite costs depending on the distance between the respective locations.

6.2 Run Time Performance

We compare the computational performance of the proposed Benders decomposition method to the IP formulation solved by CPLEX. We analyze the run-time behavior of the two main complexity drivers: the number of demands and the number of services. A run-time limit of 24 hours was imposed for all instances. The tables indicate the average run times in CPU seconds and the time improvement of the Benders decomposition compared to CPLEX, as well as the number of instances solved within the time limit by

CPLEX and the Benders decomposition (in squared brackets).

$\|\mathcal{D}\|$	CPLEX		BD		time improvement
	time (sec)	[solved]	time (sec)	[solved]	(in %)
150	11455	[22/24]	1829	[24/24]	84.03
160	19402	[20/24]	2430	[24/24]	87.48
170	29318	[18/24]	7297	[24/24]	75.11
180	35304	[18/24]	6459	[24/24]	81.70

Table 1: Run Time Depending on Number of Demands

Table 1 shows the average run time for 24 instances, four networks and six demand scenarios of the considered demand size. CPLEX could only solve 78 of the 96 instances within 24 hours to optimality, while the proposed Benders decomposition solved all instances. For the unsolved instances in CPLEX, several instances had to be stopped because of memory problems and were solved on a larger machine. The optimality gap for the unsolved instances was between 0.15% and 0.01%. The proposed solution method gives a run time improvement of more than 80% on average.

$\|\mathcal{R}\|$	IP		BD		time improvement
	time (sec)	[solved]	time (sec)	[solved]	(in %)
base case	11455	[22/24]	1829	[24/24]	84.03
base case +10	8296	[22/24]	2061	[24/24]	75.16
base case +20	20770	[20/24]	2753	[24/24]	86.74
base case +30	33609	[16/24]	4383	[24/24]	86.96

Table 2: Run Time Depending on Number of Services for $|\mathcal{D}| = 150$

We also analyzed the effect of increasing the number of services. For that, we took the six 150-demand scenarios and the four generated networks as the base case. We then added 10, 20, and 30 services to the base case. The results displayed in Table 2 yield conclusions similar to those of the previous experiments. CPLEX could only solve 80 of the 96 instances to optimality within 24 hours (the same memory issues were also observed), with optimality gaps between 0.22% and 0.01% for the unsolved ones. We note that the number of solved instances decreases when the number of services increases. With respect to the Benders decomposition, we observe the classical behavior of the run time increasing with the number of services, i.e., with the number of design variables in the master, which increase its computational complexity.

The general conclusion is that the solution method we propose outperforms CPLEX for the instances and dimensions tested. Yet, to further test the limits of our method, we increased the number of demands and the services. The results are displayed in Tables 3 and 4, respectively.

We increased the demand up to 250 customers, and observed that, on the one hand,

$\lvert\mathcal{D}\rvert$	time (sec)	[solved]	gap of unsolved (in %)
200	16023	[22/24]	2.97
220	29070	[19/24]	2.04
250	42855	[12/24]	1.99

Table 3: Run Time BD Depending on Number of Demands (larger instances)

$\lvert\mathcal{R}\rvert$	time (sec)	[solved]	gap of unsolved (in %)
~ 140	8044	[24/24]	-
~ 160	10508	[23/24]	1.44

Table 4: Run Time BD Depending on Number of Services (larger instances)

the number of instances solved to optimality decreases with the demand size, 22 out of 24 for 200 customers and 12 out of 24 for 250, but, on the other hand, many of the unsolved instances are close to the optimum as the average optimality gap decreases from about 3% to 2%. Similarly, we doubled the number of services of the basic scenario (with $|R| \sim 70$), resulting in 138 - 146 services (shown in Table 4 as approximately 140 services). Then, we further added 20 services to the network. The results show that the proposed method solves the instances with twice the number of services to optimality. One also observes that, even for the largest network, 23 out of 24 instances were solved to optimality, and that the unsolved instances ended with a gap of 1.44%.

To sum up, these experiments indicate that the method we propose outperforms CPLEX. Moreover, when the problem dimensions are increased, it either still solves many of the instances to optimality, or, when the method is stopped, the final gap is small. One also observes that the number of demands seems to impact the computational performance of the method more than the number of services. This is in line with the general observation regarding the impact of the number of commodities (customers) on the difficulty to address network design instances.

6.3 Benefits of a Multimodal Network

We used the proposed BD algorithm to analyze the benefits of the different transportation modes, and the impact of different shares of inbound and outbound demands (next subsection). For both analyzes, we considered the six 150-customer cases with inbound and outbound shares of 50%, 66% and 83%, and network 1 with four satellites. Experiments were run considering either both modes (truck and tramway) or only one.

To evaluate and compare the different settings, the following key performance indicators are used: *total cost* of the final solution; number of operated/selected services (*# services*); the average vehicle load with respect to its capacity ("utilization") over all services after being loaded at the starting CDC (*start util. (in %)*); the average vehicle load with respect to capacity when arriving at the final CDC (*end util. (in %)*); *empty km* defined as the total number of kilometers driven without freight in any compartment and *full km* defined as the total number of kilometers driven with a full vehicle (we assume a truck or a tramway is full when more than 80% of its capacity is used in a segment). For the multimodal fleet case, the results are also shown for each mode in squared brackets [tramway/truck].

modes	total costs	# services	start util. (in %)	end util. (in %)	empty km	full (80%) km
tram + truck	22624	5.7 [3.7/2.0]	54 [66/44]	28 [38/18]	125 [13/112]	122 [104/18]
tram	23464	6.3	56	29	39	109
truck	24457	6.2	41	21	328	32

Table 5: Analysis of the Impact of the Fleet Structure

Table 5 compares the performance of the system with a multimodal fleet and the two single-mode cases. One observes that the multimodal fleet structure significantly reduces the costs and the number of operated services, it also improves the vehicle utilization. Due to multiple compartments, which allow combining inbound and outbound demands, the tramway is well used and almost never empty. This is true for all the cases when the line-based services are part of the system. Note that the higher vehicle-capacity utilization out of the initial CDC of the service, compared to when arriving to the final CDC, is normal given that more demand is entering the city than leaving it.

We noticed that, in general, trucks were not loaded as tightly as tramways. It appears that in these experiments trucks were loaded to deliver at the first satellite and then continued to load outbound demand at a second satellite. This reflects the combination of the need to completely unload the truck before being able to use it to load outbound goods, and the limited satellite vehicle-handling capacity at each period. Tramways, on the other hand, benefit from the multiple compartments. They need to empty one compartment only before being used for outbound traffic, and they can therefore use the satellites for inbound and outbound demand simultaneously. Nevertheless, the results also indicate that a single-mode tramway system is also not preferable, as its total service travel time is typically longer than that of the truck due to longer routes (back and forth on the line) of tramways. A multimodal 2T-CL thus appears the beneficial.

6.4 Benefits of Combining Inbound and Outbound Demands

We used the same setup as in the previous section to examine the benefits of combining inbound and outbound demands (Fontaine et al., 2017, discuss the results of a preliminary analysis of this characteristic without considering transportation modes and resources).

We compare the results of our full model, with the combined inbound and outbound demands, with the results of solving an inbound-demand model and an outbound-demand model separately. Considering the previously defined key performance indicators, Table 6 displays the relative (%) improvement of combining both demand types in one model as apposed to modeling them in separation. The results are aggregated according to the three levels of inbound share out of the total demand, as well as the three multimodal and unimodal cases. Vehicle utilization is compared separately for inbound and outbound services.

$\lvert\mathcal{D}^I\rvert/\lvert\mathcal{D}\rvert$	modes	total costs	# services	start util.	end util.	empty km	full km (80 %)
0.5	tram + truck	-15.61	-33.33	6.19	-45.69	-82.57	-63.87
	tram	-24.17	-33.33	-14.98	-50.00	-100.00	-78.20
	truck	-12.21	-31.25	0.00	-55.91	-62.37	-74.84
0.67	tram + truck	-12.73	-25.00	6.34	-56.95	-84.72	-14.92
	tram	-17.14	-27.78	0.00	-62.93	-86.73	-50.60
	truck	-8.23	-22.22	0.00	-71.43	-52.19	-100.00
0.83	tram + truck	-13.08	-25.00	0.00	-61.50	-71.44	-5.66
	tram	-16.00	-27.78	0.00	-63.16	-93.28	-18.07
	truck	-10.27	-25.00	0.00	-66.67	-42.88	-41.89

Table 6: Improvement when Combining Inbound and Outbound Demands (in %)

The results show that the total cost and the number of services are always reduced. Not surprisingly, reductions are more substantial when the inbound and outbound demands are more balanced, the 0.5 ratio in our case. However, improvements are still significant as the ratio increases, and these improvements appear fairly stable even for high ratio values. This observation is important as most cities display at least a 20% share of outbound demand. A second observation concerns the vehicle utilization. While the number of empty kilometers is reduced, the number of full kilometers is reduced as well. The latter is due to the fact that, combining inbound and outbound flows for trucks often implies that vehicles are "rapidly" emptied of the inbound demand, to allow for loading outbound demands, resulting in relatively low volumes being loaded.

Table 6 further indicates that the benefits for (almost) all key performance indicators are strongest in the single-mode tramway case and the weakest in the single-mode truck setting. This underlines the benefits of a multiple compartment, multimode structure for multiple and efficient utilization of the vehicles. The results of the truck utilization are not surprising and are in line with those of the load-distribution analysis of the previous section. When trucks are the only mode, particularly in the single outbound demand case, fewer and better loaded services are used. In the mix-demand case, outbound loads are put on the trucks that brought inbound loads to the satellites. This is good from a cost point of view, but not necessarily from an efficient-loading point of view. This indicates that performance measures related to vehicle utilization, in particular the number of full and empty kilometers traveled, have to be treated carefully when evaluating such a system.

To conclude, we can state that the integration of inbound and outbound demand in the same model should be considered for an efficient City Logistics system and resource utilization.

7 Conclusions and Outlook

In this paper, we addressed the tactical planning of an extended two-tier City Logistics system integrating inbound and outbound demand, different transportation modes combining traditional, road-based, carriers and massive-flow modes and vehicles such as light and regular rail. assignment of customers to consolidation distribution centers and satellites, and the management of a number of major resources, such as the multiple facets of satellite capacity and the structure, allocation, and size of the heterogeneous fleets. According to our best knowledge, these systems and the associated planning problems were to addressed in the literature. We proposed a scheduled service network design formulation for the tactical planning of such extended systems, as well as an efficient Benders decomposition algorithm, which includes a tailored partial decomposition technique for deterministic mixed-integer linear-programming formulations, and several valid inequalities and pareto-optimal cuts. The results of extensive numerical experiments show the efficiency of the proposed solution method, as well as the benefits to the flexibility and efficiency of a city logistics network of integrating several demand types and multimodal transportation networks into a single formulation.

To conclude, we mention two challenging research directions. The first concerns solution methods for larger problem settings. Matheuristics appear promising as one could benefit from the well-studied knapsack problem results. Considering the uncertainty in the problem parameters makes up a second interesting research direction.

Acknowledgments

While working on the project, the first author was postdoctoral fellow at the School of Management, Université du Québec à Montréal and member of the Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), while the second author was Adjunct Professor, Department of Computer Science and Operations Research, Université de Montréal. We gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery and Acceleration Grant programs. We also gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants, and their new university researchers start up program. Thanks are also due to Calcul Québec and Compute Canada for providing the authors access to their high-performance computing infrastructure.

References

N. Arvidsson and M. Browne. A review of the success and failure of tram systems to carry urban freight: the implications for a low emission intermodal solution using electric vehicles on trams. Technical report, EUT Edizioni, Università di Trieste, 2013.

J. Barceló, H. Grzybowska, and S. Pardo. Vehicle Routing and Scheduling Models, Simulation and City Logistics. In Zeimpekis, V., Tarantilis, C.D., Giaglis, G.M., and Minis, I., editors, *Dynamic Fleet Management Concepts - Systems, Algorithms & Case Studies*, pages 163–195. Springer, 2007.

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

J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik*, 4(1):238–252, 1962.

A. Benjelloun and T.G. Crainic. Trends, Challenges, and Perspectives in City Logistics. *Buletinul AGIR*, 4:45–51, 2009.

M. Boccia, T.G. Crainic, A. Sforza, and C. Sterle. A Metaheuristic for a Two-Echelon Location-Routing Problem. In P. Festa, editor, *Experimental Algorithms*, volume 6049 of *Lecture Notes in Computer Science / Programming and Software*, pages 288–301. Springer-Verlag, Berlin Heidelberg, 2010.

M. Boccia, T. G. Crainic, A. Sforza, and C. Sterle. Multi-Commodity Location-Routing: Flow Intercepting Formulation and Branch-and-Cut Algorithm. *Computers & O.R.*, 89:94–112, 2017.

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

M. Brandt. Frankfurt: "logistiktram". *European Review of Regional Logistics*, 4:5, 2018.

U. Breunig, V. Schmid, H. R.F., and T. Vidal. A Large Neighbourhood Based Heuristic for Two-Echelon Routing Problems. *Computers & Operations Research*, 76:208–225, 2016.

C. Contardo, V.C. Hemmelmayr, and T.G. Crainic. Lower and Upper Bounds for the Two-Echelon Capacitated Vehicle Routing Problem. *Computers & O.R.*, 39(12):3185–3199, 2012.

T. G. Crainic. Service network design in freight transportation. *European Journal of Operational Research*, 122(2):272–288, 2000.

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

T. G. Crainic and B. Montreuil. Physical internet enabled hyperconnected city logistics. *Transportation Research Procedia*, 12:383–398, 2016.

T. G. Crainic and A. Sgalambro. Service network design models for two-tier city logistics. *Optimization Letters*, 8(4):1375–1387, 2014.

T. G. Crainic, N. Ricciardi, and G. Storchi. Advanced freight transportation systems for congested urban areas. *Transportation Research Part C: Emerging Technologies*, 12 (2):119–137, 2004.

T. G. Crainic, N. Ricciardi, and G. Storchi. Models for evaluating and planning city logistics systems. *Transportation science*, 43(4):432–454, 2009.

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

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

T. G. Crainic, M. Hewitt, F. Maggioni, and W. Rei. Partial benders decomposition strategies for two-stage stochastic integer programs. Publication CIRRELT-2016-37, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Université de Montréal, Montréal, QC, Canada, 2016b.

T. G. Crainic, M. Hewitt, M. Toulouse, and D.M. Vu. Scheduled Service Network Design with Resource Acquisition and Management. *EURO Journal on Transportation and Logistics*, 2017.

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

T.G. Crainic. City logistics: Advanced urban freight transportation systems. In Z.-L. Chen and S. Raghavan, editors, *Tutorials in Operations Research 2008, State-of-the-Art Decision Making Tools in the Information-Intensive Age*, Series on Computers and Operations Research, pages 181–212. INFORMS, 2008.

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

T.G. Crainic, F. Errico, W. Rei, and N. Ricciardi. Integrating c2e and c2c Traffic into City Logistics Planning. *Procedia - Social and Behavioral Sciences*, 39:47–60, 2012.

J. Danard and K. Janin. Tramfret project. `http://tramfret.com/`, 2016. Accessed: 2018-09-27.

M. Dawande, J. Kalagnanam, P. Keskinocak, F. S. Salman, and R. Ravi. Approximation algorithms for the multiple knapsack problem with assignment restrictions. *Journal of Combinatorial Optimization*, 4(2):171–186, 2000.

H. Dyckhoff and U. Finke. *Cutting and packing in production and distribution: A typology and bibliography*. Springer Science & Business Media, 1992.

A. Erera, M. Hewitt, S. M.W.F., and Y. Zhang. Improved load plan design through integer programming based local search. *Transportation Science*, 47(3):412–427, 2013.

P. Fontaine, T. G. Crainic, O. Jabali, and W. Rei. The impact of combining inbound and outbound demand in city logistics systems. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pages 766–770. IEEE, 2017.

Y. Freemark. Opportunities abound for transporting goods by tram - if properly coordinated. http://www.thetransportpolitic.com/2011/10/23/opportunities-abound-for-transporting-goods-by-tram-if-properly-coordinated/, 2011. Accessed: 2017-08-17.

V. Ghilas, E. Demir, and T. van Woensel. Integrating Passenger and Freight Transportation: Model Formulation and Insights. Beta working paper 441, Research Schoolk for Operations Management and Logistics, Eindhoven Technical University, Eindhoven, The Netherlands, 2013.

P. Gianessi. *Solving Strategic and Tactical Optimization Problems in City Logisticsl*. PhD thesis, Université Paris 13, France, 2014.

P. Gianessi, L. Alfandari, L. Ltocart, and R. W. Calvo. The multicommodity-ring location routing problem. *Transportation Science*, 50(2):541–558, 2016.

V. C. Hemmelmayr, J.-F. Cordeau, and T. G. Crainic. An adaptive large neighborhood search heuristic for two-echelon vehicle routing problems arising in city logistics. *Computers & Operations Research*, 39(12):3215–3228, 2012.

M. Hewitt. Enhanced Dynamic Discretization Discovery for the Continuous-Time Service Network Design Problem. *Transportation Science*, 2018. forthcoming.

J. Holguín-Veras, C. Wang, M. Browne, S.D. Hodge, and J. Wojtowicz. The new york city off-hour delivery project: Lessons for city logistics. *Procedia - Social and Behavioral Sciences*, 125:36–48, 2014.

Y. Huang, L. Zhao, W.B. Powell, Y. Tong, and I.O. Ryzhov. Optimal Learning for Urban Delivery Fleet Allocation. *Transportation Science*, 2019. https://doi.org/10.1287/trsc.2018.0861.

G. Laporte and F. V. Louveaux. The integer l-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142, 1993.

G. Laporte, F. V. Louveaux, and L. Van Hamme. An integer l-shaped algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research*, 50 (3):415–423, 2002.

M. Lindholm and S. Behrends. Challenges in urban freight transport planning–a review in the baltic sea region. *Journal of Transport Geography*, 22:129–136, 2012.

T. L. Magnanti and R. T. Wong. Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–484, 1981.

S. Mancini, J. Gonzalez-Feliu, and T. G. Crainic. Planning and Optimization Methods for Advanced Urban Logistics Systems at Tactical Level. In J. Gonzalez-Feliu, F. Semet, and J.-L. Routhier., editor, *Sustainable Urban Logistics: Concepts, Methods and Information Systems*, pages 145–164. Springer, 2014.

R. Masson, A. Trentini, F. Lehuédé, N. Malhéné, O. Péton, and H. Tlahig. Optimization of a city logistics transportation system with mixed passengers and goods. *EURO Journal on Transportation and Logistics*, 6(1):81–109, 2017.

J. Morana, J. Gonzalez-Feliu, and F. Semet. Urban Consolidation and Logistics Pooling - Planning, Management and Scenario Assessment Issues. In J. Gonzalez-Feliu, F. Semet, and J.-L. Routhier, editors, *Sustainable Urban Logistics: Concepts, methods and Information Systems*, pages 187–210. Springer, 2014.

P.K. Nguyen, T.G. Crainic, and M. Toulouse. Multi-Trip Pickup and Delivery Problem with Time Windows and Synchronization. *Annals of Operations Research*, 253(2): 899–934, 2017.

N. Papadakos. Practical enhancements to the magnanti–wong method. *Operations Research Letters*, 36(4):444–449, 2008.

G. Perboli, R. Tadei, and D. Vigo. The two-echelon capacitated vehicle routing problem: models and math-based heuristics. *Transportation Science*, 45(3):364–380, 2011.

R. Rahmaniani, T. G. Crainic, M. Gendreau, and W. Rei. The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3): 801–817, 2017.

M. I. Restrepo, B. Gendron, and L.-M. Rousseau. Combining benders decomposition and column generation for multi-activity tour scheduling. Publication CIRRELT-2015-57, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Université de Montréal, Montréal, QC, Canada, 2015.

H. Riemann. Logistiktram project. http://logistiktram.de/, 2019. accessed: 2019-01-16.

A. Rusish. *Collaborative Logistics Networks*. PhD thesis, Università degli studi di Trieste, Italy, 2017.

M. Savelsbergh and T. Van Woensel. City logistics: Challenges and opportunities. *Transportation Science*, 50(2):579–590, 2016.

E. Taniguchi. Concepts of city logistics for sustainable and liveable cities. *Procedia-social and behavioral sciences*, 151:310–317, 2014.

E. Taniguchi, R.G. Thompson, T. Yamada, and J.H.R. van Duin. *City Logistics: Network Modelling and Intelligent Transport Systems*. Pergamon, Amsterdam, 2001.

E. Taniguchi, R.G. Thompson, and T. Yamada. Recent trends and innovations in modelling city logistics. *Procedia -Social and Behavioral Sciences*, 125:4–14, 2014.

A. Trentini and N. Mahléné. Toward a shared urban transport system ensuring passengers & goods cohabitation. *Tema. Journal of Land Use, Mobility and Environment*, 3(2): 37–46, 2010.

V. Tsurkov. *Large-scale Optimization: Problems and Methods*, volume 51. Springer Science & Business Media, 2013.

M. Winkenbach, R.S. Kleindorfer, and S. Spinler. Enabling urban logistics services at la poste through multi-echelon location-routing. *Transportation Science*, 50(2):520–540, 2016.

8 Annexes

8.1 Notation

Notation	Description
$t = 1, \dots, T$	planning horizon
$\mathcal{M}$	set of means of transport
$\mathcal{E}$	set of city distribution centers
$\mathcal{Z}$	set of satellites
$\mathcal{Z}(d)$	set of feasible satellites for demand d
$\mathcal{T}$	set of urban-vehicle types
$\mathcal{D}$	set of demands
$\mathcal{D}^I$, $\mathcal{D}^O$	set of inbound, outbound demands
$\mathcal{R}$	set of urban-vehicle services
$\mathcal{R}(z,t)$	set of urban-vehicle services, present at satellite z at time t
$\mathcal{R}(z,t,m)$	set of urban-vehicle services of mode m, present at satellite z at time t
$\mathcal{R}(t,\tau,e)$	set of urban-vehicle services of type τ, starting at external zone e and operating during time t
$\mathcal{R}^C(r)$	set of compartment services of urban-vehicle service r
$n_{e\tau}$	fleet size of urban-vehicle τ at city distribution center e
u^c_τ	compartment capacity of urban-vehicle τ
a_{zt} [a^m_{zt}]	urban-vehicle limit at satellite z in period t [for mode m]
b_{zt}	loaded/unloaded volume limit at satellite z in period t
v_d	volume of demand d
e_r	city distribution center of urban-vehicle service r
τ_r	urban-vehicle type of urban-vehicle service r
m_r	mode of of urban-vehicle service r
k_r	operating costs of urban-vehicle service r
f_{de}	costs for assigning demand d to city distribution center e
s_{dzr}	costs for assigning demand d to satellite z for service r

Table 7: Notation

8.2 Networks

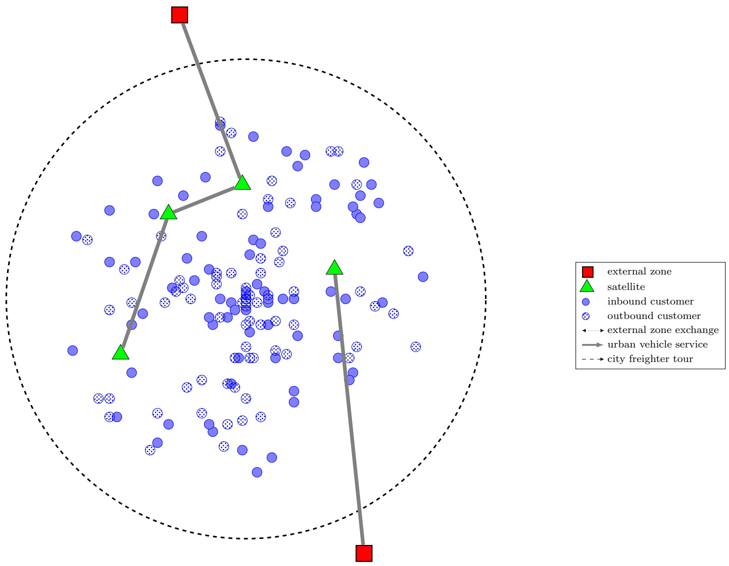

Figure 3: Network 1

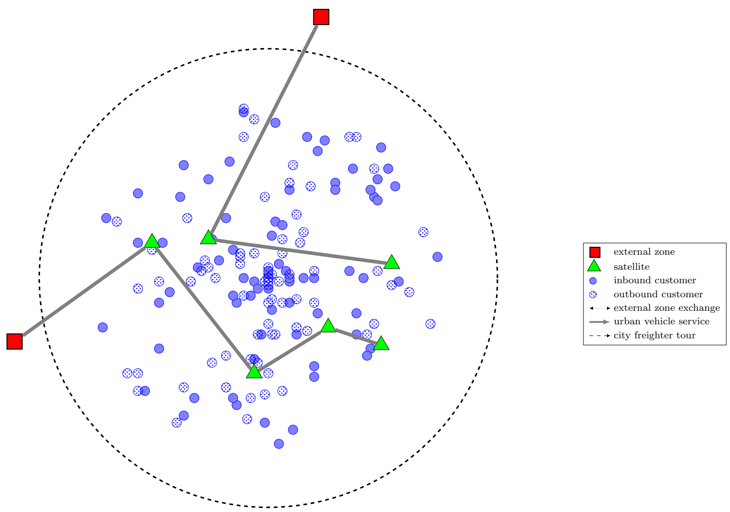

Figure 4: Network 2

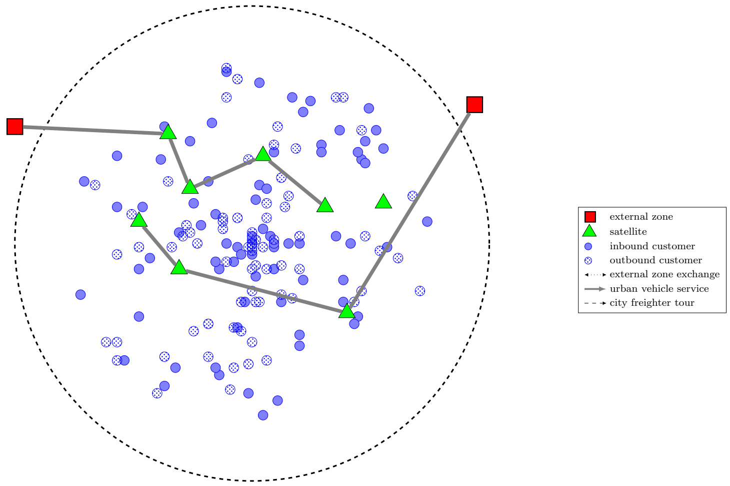

Figure 5: Network 3

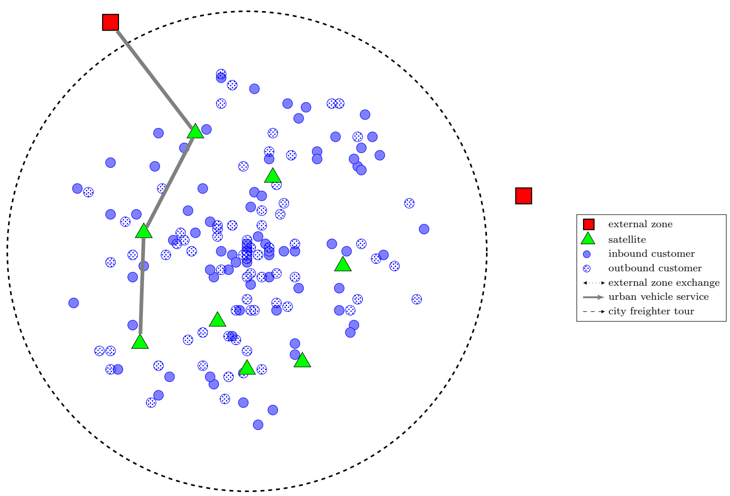

Figure 6: Network 4