Supply Chain Logistics & Methods
TA4: Humanitarian Logistics
Thursday 8:30 – 10:30 AM
Session Chair: Lavanya Marla

8:30  Using Drones to Minimize Latency in Distribution Systems
Mohammad Moshref-Javadi*, Seokcheon Lee
Purdue University

9:00  A Novel Formulation and a Column Generation Technique for a Rich Humanitarian Logistic Problem
Ohad Eisenhandler, Michal Tzur*
Tel Aviv University

9:30  Humanitarian Medical Supply Chain in Disaster Response: Role and Challenges
1Irina Dolinskaya*, +2Maria Besiou, 2Sara Guerrero-Garcia
1Northwestern University, 2Kühne Logistics University

10:00 Cooperative Humanitarian Logistics Models for Highly Resource-Constrained Settings
Lavanya Marla*
University of Illinois at Urbana-Champaign
Using Drones to Minimize Latency in Distribution Systems

Mohammad Moshref-Javadi, Seokcheon Lee

School of Industrial Engineering, Purdue University, 315 N. Grant St., West Lafayette, IN 47907, USA
moshref@purdue.edu
lee46@purdue.edu

Abstract

Use of drones for the purpose of delivery of service and goods has recently gained considerable attention. Drones are specifically suitable for the delivery of small, urgent, and light packages, for example, blood and medicine in emergency situations and disaster relief. In this research, we consider a combined delivery system of drones and trucks for more efficient delivery of products. It is assumed that a single truck carries both goods and drones and launches the drones at some points in its route. At each stop point, the truck waits until all drones come back to the truck and then it can move to the next customer. We also extend this system and assume that the drones may be launched multiple times when the truck stops. The results show that using drones can considerably improve the waiting time of recipients. Comparison of the single-trip with multi-trip drone systems shows that having multiple trips of drones at each truck’s stop is very effective in reducing the latency at customers.

Extended Abstract

In customer-oriented distribution systems, the focus is on minimization of the waiting time of the customers in order to maximize customer satisfaction which finally leads to increased profit. Effective design of the last mile delivery in customer-oriented systems can help reduce this waiting time as well. In addition to the customer-oriented commercial systems, reducing the waiting time of victims plays an important role in post-disaster relief logistics. After a disaster, victims need both supplies and services (aids). The urgency of the help is usually the most important factor because the rate of loss and death will increase with respect to the latency of delivery. This latency is directly affected by delivery, i.e., more people will survive with less suffering by faster delivery.

Recently the applications of drones have been proposed for more efficient last-mile delivery (Murray and Chu 2015) (Ponza 2016) (Agatz, Bouman, and Schmidt 2015) (Ha et al. 2015) (Olivares et al. 2015) (Savuran and Karakaya 2015) (Ferrandez et al. 2016) (Wang, Du, and Ma 2014). In this approach, trucks carry some drones in addition to packages. At some points in their routes, drones are launched to deliver supplies to customers. Thus, not only trucks, but also drones are used to deliver supplies to recipients. One of the main reasons to use drones is their faster speed and no need for roads. However, due to capacity constraints and flight range limits, they are not able to be used as the sole vehicle for distributing supplies. The combination of trucks and drones seem to be an effective approach for last-mile delivery. In this combined system, trucks are used for two purposes: deliver larger packages, carry drones to a location close to the customers’ locations and launch them for faster delivery.
Because drones can fly and they are not confined to roads and consequently traffic, their speed of delivery is higher. They can travel in three dimensions space from and to any directions. Therefore, due to no traffic, faster delivery is one of the main and appealing reasons of using drones. However, higher speed is not the only reason of popularity of drone delivery. Drones do not need any roads for travel like cars. Therefore, they can fly in areas in which there is not suitable road system, such as rural areas and islands. Matternet has been one the pioneers in drone delivery which used drones for distribution of crucial supplies in Haiti, South Africa after the earthquake (Matternet First Delivery). They also used drones in Lesotho, the capital, for delivering blood samples from clinics to hospitals to be analyzed for HIV/AIDS (Lesotho First Delivery). Lesotho does not have a well-paved road network and drones can considerably help the delivery of products. In fact, the delivery of blood samples and medicines are perfect products to be distributed by drones because of size, weight, value, and urgency. In addition to road networks, even after a flood disaster, delivery by drones can be a convenient method since most of the roads are covered with water after a flood. Delft University of Technology (TU Delft) also launched a video about ambulance drones that can help patients in emergency situations.

In spite of many advantages such as faster delivery because of no traffic and no need for roads, and environmentally friendly aspects, drones have some limitations. Currently, the main limitation for drone delivery is the Federal Aviation Administration (FAA) regulations which only allows drones to travel in a specific distance range where the drones can be seen in sight. However, one of the main limitations of drones is package size, weight, and travel limitation. Drones are able to carry small packages, with respect to both size and weight. Thus, drones have very high marginal costs for packages with a little greater size or weight. Also, due to battery limitations, they can travel in a limited range of continues travel, approximately 30 minutes. Nonetheless, according to a study by Amazon (Amazon Drone Analysis), 86% of its packages weigh under 5 pounds and 70% of Americans live within 5 miles of a Walmart. Thus, even if we suppose that drone synchronization with trucks is impossible, still drones have a good chance and market to be used for delivery directly from stores to customers for most of the product sizes and weights.

In this research, we propose the applications of drones in distribution operations by taking into account the advantages of faster speed, no need for roads, and no traffic. According to these significant advantages, drones are suitable for deliveries with the goal of minimum waiting time of recipients. Thus, two main applications with respect to this objective are: first the delivery of packages and services in emergency situations similar to drone ambulances (TU Delft) or aid distribution after disaster. The second application is in commercial systems in which customer satisfaction is one of the main goal, which will finally lead to more profit obtained from satisfied customers.

The routing problem which focuses on the waiting time of the customers is called the Traveling Repairman Problem (TRP) (Blum et al. 1994) in which a vehicle distribute products to customers with the goal of minimum sum of latencies at customers. In this research, drones are incorporated in the Traveling Repairman Problem. Therefore, we propose a combined problem, called the Traveling Repairman Problem with Drones (TRPD). In this problem, a truck is assumed to deliver products to customers. This truck also carries some drones and launch them at some points in its route to deliver products to customers. Thus, some of the customers receive packages by the truck while the rest of the customers are visited by drones. The goal of the problem is minimization of the waiting time of among all customers. This goal is defined as two different objective functions in the proposed formulations. The first objective function minimizes the sum of latencies at customers, while the second objective function minimize the largest latency, i.e., the waiting time of the last visited customer, either by the truck or drones.
A worst case analysis of the problem will be conducted and compares the Traveling Salesman Problem with Drones and TRP. One of the worst case analyses presents that using \( k \) drones can lead to improvement of the objective function maximum latency with a factor of \( 2k+1 \). That is, if the maximum latency is 100 in the TRP, adding 2 drones to the problem and solving the TRPD can lead to objective function value 20, which is a significant reduction in the latency. Other results are also related to the comparison of the TRPD with the Vehicle Routing Problem with Drones and comparison of TRPD with TRP on the objective function sum of latencies.

The problem is formulated as mixed integer programming models. In this formulation it is assumed that when a truck stops to launch vehicles, each drone can take a single trip to a customer and return to the vehicle. This drone will not be relaunched until the vehicle moves from the current customer to another customer. Thus, it has to wait until all drones return to the vehicle and then the truck can continue the route. The truck is also able to serve customers on its way while drones are in the truck. We assume a single vehicle distributes products to customers from a single depot. This vehicle carries both products and multiple drones which will be launched to deliver products. Speed of the drones can be higher than the speed of the vehicle. Each drones can serve one customer in each visit. It is also assumed that the vehicles can only stop at customer nodes to launch drones. In other words, the network that the vehicle can stop in is not continuous.

Figure 1 illustrates the problem schematically.

![Figure 1: TRPD with single-trip for drones](image)

In this model, if a vehicle launches some drones and the travel times of the drones to customers differ considerably, some of the drones are done with their deliveries quickly and has to wait for other drones to return to the vehicle. This type of delivery seem not to be very efficient. Thus, to resolve this issue and add more flexibility to our model, we assume that drones can take multiple trips to deliver packages to customers whenever the vehicle stops at a customer. Thus, when drones come back to the vehicle after the first delivery, they may pick-up another package and start a new delivery while the vehicle still stops at the same node. Thus, some drones can have multiple small trips, while the rest of the drones may have some long trips.
The mathematical models are used to solve several problems with different configuration. The network of customers have two different configuration of random uniform and random clustered. The location of depot may be the (0,0) or centroid. 10 problem instances are generated for each configuration. The mathematical models are also used to solve a case study in Richmond, VA with 9 customers, 2 or 3 drones, and three values for flight range of drones.

The results of the case study show that adding 2 and 3 drones reduce the latency by 38% and 58%, respectively. Also, assuming multi-trip for drones is very effective in reducing latency and can improve the results by 50%. On the other hand, the limited flight range of drones can increase the latency more than 43%. Therefore, battery limitation and FAA regulations about line of sight should be taken in to consideration when these kinds of drones systems are designed.

References:


TU Delft: https://www.youtube.com/watch?v=y-rEl4bezWc

Lethoso First Delivery: https://www.flexport.com/blog/drone-delivery-economics/
Amazon Drone Analysis: https://www.flexport.com/blog/drone-delivery-economics/
A Novel Formulation and a Hybrid Solution Procedure for a Rich Humanitarian Logistic Problem

Ohad Eisenhandler, Michal Tzur

Department of Industrial Engineering, Tel Aviv University, Tel Aviv, Israel

1. Introduction

Food rescue is the collection of perishable products from food suppliers who are willing to make donations, and its distribution to welfare agencies, such as food pantries, soup kitchens, homeless shelters, orphanages and schools, that serve individuals in need. These operations are managed by food banks, which are nonprofit charitable organizations. Their operations have become exceedingly widespread in recent years due to economic crises that have increased the demand for nutritional aid, and the benefit to donors who can avoid in this way the costs of destroying excess production while reflecting a social-aware image.

In this work we address a problem which is inspired by the daily activity of food banks in Israel (the "Latet" organization) and in the US (the Houston Food Bank) and specifically by the logistic challenges they encounter. First, they have to decide which supplier sites (where donations are collected) and agencies (where food is dropped off) should be visited and how to integrate them into vehicle routes. Each route cannot exceed the driver's work hour limitation, and therefore not all suppliers and agencies may be visited. In addition, they must plan how much food should be picked up from each supplier and how much of it should be delivered to each agency. A major consideration is how to maintain equity towards the different agencies, in each period, over time and over product mix, while delivering as much food as possible in total. These decisions need to be made simultaneously with routing decisions. This problem was introduced and modeled in Eisenhandler and Tzur (2016), who refer to it as The Humanitarian Pickup and Distribution Problem (H-PDSP). While the problem resembles in its structure several families of well-studied routing problems, such as pickup and delivery problems, inventory routing problems and routing problems with profits, it differs from them in its "humanitarian" objective function. In the growing stream of literature on food aid, simultaneous routing and resource allocation has been considered in only very few papers, e.g., Lien et al. (2014) and Balcik et al. (2014).

2. Preliminaries

As mentioned, this study builds on the analysis presented in Eisenhandler and Tzur (2016), where a basic, static variant of the H-PDSP was studied. We briefly provide the key concepts which are relevant to the current study. A set of pickup sites (food suppliers) and a set of
delivery sites (welfare agencies) are both given. Each supplier may announce a donation of a given size, before the vehicle leaves the depot. Each welfare agency serves a given number of individuals and has no initial inventory. We assume each site can be visited at most once, and that there is no significance to the time of the day when the donations are delivered to the agencies. A traveling time matrix is given, as well as loading or unloading times at each site (we assume that these times are constant and depend only on the site where they occur). There is a limitation on the capacity of the vehicle used for the distribution of the food, as well as a time limitation on the duration of the operation. We also assume that the quantities to be picked up or delivered are not required to be integral, and that a quantity of any size is meaningful. The problem is to decide which pickup (delivery) sites to visit, how much to pick up (deliver) from (to) each site, and in what order to visit the sites.

A major issue of the study was on designing an objective function to balance two considerations: (a) the effectiveness of the operation, measured as the total amount of units supplied to all agencies; and (b) its equity, the extent to which it maintains fair allocations among the agencies, measured with the Gini Coefficient. Based on this new objective function, a mathematical model for the problem, which we refer to as the site-based formulation, was presented. Unsurprisingly, the problem was shown to be NP-Hard under this new objective function, and this was evident in the results of the numerical experiments as well, since larger instances obtained high optimality gaps after one-hour runs. However, it was shown that when the route is pre-determined, the remaining allocation sub-problem (ASP) can be solved efficiently with a dedicated procedure, known as the Robin Hood (RH) algorithm. A key observation was that the problem can be reduced to segment-form, where a segment is a sequence of sites, having the pickup before the delivery sites. Each such segment has an initial wealth, which is the amount allocated per individual, i.e., the total supply of the pickup sites it includes divided by the total population of the delivery sites it includes. Optimality conditions which were proved for the ASP stated what should be the pickup amount in each segment, and under what conditions transfers of the initial wealth should be performed between consecutive segments. When a transfer is performed, the vehicle should carry a positive load when moving from a certain segment to its successor in the route.

The RH algorithm plays a key role in a heuristic solution method developed for the problem, based on the Large Neighborhood Search (LNS) framework. The LNS algorithm consists of addressing the routing sub-problem using a local search over “broad” neighborhoods, and each route considered is evaluated by solving its corresponding allocation sub-problem. This method was shown to perform well in previous parts of the research.
3. A New Formulation

In this work, we use the segment-form structure of the solution, along with the RH algorithm, as a means to present a new mathematical formulation of the problem. We substitute the "classical" site-based routing decision variables, i.e., whether the vehicle should proceed from a certain site to another, with new variables which indicate whether the vehicle should proceed from a certain sequence of sites (which are also a sequence of segments) to another. These sequences are chosen in a way that guarantees that the allocation decisions which they dictate, can be made independently of the other sequences that are chosen in the solution.

The main contributions of this new formulation are in the following directions:

1. Novelty: We use a new modeling approach, which has not been used previously in the literature, to the best of our knowledge. It is inherently different from other known formulations for routing problems which include exponentially many decision variables, e.g., the set partitioning formulation of the CVRP (covered in Semet et al., 2014) or the set packing formulation of the Team Orienteering problem (e.g., Boussier et al., 2007).

2. Strength: The linear relaxation of this new formulation can be shown to provide a tighter bound than the linear relaxation of the site-based formulation used in previous work, see Theorem 1 below.

3. Applicability: This formulation gives rise to a new solution methodology for the problem, based on a hybrid approach.

The main idea of the new formulation is based on the following definition. A super-segment is defined as a sequence of segments, having non-increasing wealths. We note that a single segment can also be considered as a super-segment according to this definition. A complete vehicle route is then a sequence of super-segments which are mutually exclusive in the sites they include. Note that the definition of super-segments implies that there is no need for transfers between consecutive super-segments in a route. Thus, the allocation of each super-segment can be computed independently, and it suffices to determine the sequence of super-segments the vehicle visits. Note that the vehicle can move directly from one super-segment to another only if: (1) these super-segments are mutually exclusive in the sites they include; (2) visiting both of them does not violate the time constraint; and (3) the wealth of the last segment of the first super-segment is strictly lower than the wealth of the first segment of the second super-segment.

As indicated above, a major advantage of the super-segment-based formulation over the site-based formulation lies in the fact that its linear relaxation provides a tighter upper bound. **Theorem 1:** Any feasible solution to the linear relaxation of the super-segment-based formulation can be transformed to a feasible solution to the linear relaxation of the site-based formulation, having the same objective value, but not vice versa.
Due to the space limitation, we provide only the general lines of the proof. To see the first direction, consider that given a feasible solution to the linear relaxation of the super-segment-based formulation, one can easily construct an equivalent routing solution to the linear relaxation of the site-based formulation. To do this, we use the fractional flows between the super-segments, as the fractional flows between the sites included in them. If the fractional flow between any pair of super-segments is included in more than one super-segment or between more than one pair of super-segments, the site-based formulation is defined to be the sum of all of these fractional flows. Once the routing solution is determined in this manner, determining the remaining allocation decisions is the equivalent of solving instances of the ASP. By showing that the allocation solution of the super-segment-based formulation satisfies the optimality conditions of these ASPs, we obtain a feasible solution to the site-based formulation having the same objective value.

To see that the other direction does not hold, note that fractional subtours between delivery sites may occur in the fractional solution of the site-based formulation (even when the relevant subtour elimination constraints are included in the formulation). Solutions including such subtours necessarily cannot be represented as a fractional solution to the super-segment-based formulation since such a sequence of delivery sites should be included, by definition, in the same super-segment. However, the formulation does not consider super-segments in which a site is repeated more than once.

4. **Hybrid Solution Methodology**

As mentioned, the main challenge of this formulation lies in the fact that it includes an exponential number of binary decision variables and constraints. To overcome this, a column generation approach w.r.t. the routing variables would typically be used in order to solve the linear relaxation of the problem at hand. In this framework, only an initial subset of columns is included in the formulation, and a pricing sub-problem is iteratively solved in order to identify new variables that should be added to the formulation. However, in the case of the H-PDSP, the contribution of each variable to the objective value is not separable in the sites that are included in its super-segments. This makes the underlying pricing sub-problem hard to solve with good theoretical or at least practical performance.

Instead, we use a different approach, in which a subset of “good” segments is used as a building block for “good” super-segments. These segments are found through short runs of the basic LNS heuristic. The improving solutions found during these runs are broken down to their constituent segments, which are then used to create super-segments for the mathematical formulation. The formulation is then solved, and its solution is used as a starting point for the next run of the LNS heuristic. This process is iteratively repeated until stopping criteria are
satisfied. Numerical experiments suggest that this approach obtains better solutions compared to the basic LNS heuristic.

References


HUMANITARIAN MEDICAL SUPPLY CHAIN IN DISASTER RESPONSE: ROLE AND CHALLENGES

Irina Dolinskaya, Maria Besiou, Sara Guerrero-Garcia

INTRODUCTION

In a large humanitarian disaster such as earthquake, hurricane, flood and others, the humanitarian response consists of multiple stages and facets of operation (e.g., initial assessment, search and rescue, food and water distribution). The existing academic literature primarily focuses on the distribution of various supplies needed by the affected population. However, little work is done specifically studying the humanitarian medical supply chain (HMSC) aspect of the. In this paper we close this gap by (1) describing the humanitarian medical supply chain in the case of disaster response, and (2) identifying the factors affecting its effectiveness, especially focusing on the factors that are unique to the medical aspect of the humanitarian supply chain.

The focus of our work is the medical response to a large disaster (initiated by a non-medical event) where the medical aspect of the response is secondary or as a result of the larger humanitarian disaster, such as earthquake, flood, etc. In such cases, medical response is only part of the relief operation. We intentionally distinguish disaster from a large scale medical emergency or medical crisis (e.g., wide spread of an infectious disease, such as Ebola outbreak in 2014), where the medical aspect of the response is the primary operation. For example, the medical response to the Haiti earthquake was crucial due to the large number of injured people, however it was of secondary role in comparison to the primary role of the medical response to the cholera outbreak that followed. We observe that the existing literature studying supply chain in a post disaster setting primarily focuses on generic aspects of the response, overlooking the unique characteristics of a medical supply chain. On the other hand, the academic literature that studies medical crises concentrates on the medical supply chain that is developed and operated mainly by the public sector (for example government) associated with setting up and supporting healthcare infrastructure, such as temporary hospitals. In this paper, we contribute to the former work by studying the HMSC and the corresponding factors that affect its effectiveness in a disaster response setting where the government’s role is not so strong and humanitarian organizations are obliged to take the lead. While the main focus of our work is on post disaster setting of non-medical disasters,
we also make a contribution to the latter body of literature by identifying the factors that affect the medical supply chain in the case of a response to a medical disaster, Ebola outbreak in western Africa in 2014.

The contributions of this paper are as follows. (1) We describe the humanitarian medical supply chain, including different stages of the supply chain and the stakeholders that play role in the HMSC. (2) Based on the academic and practitioner literature, we identify the factors affecting the HMSC in case of disaster response. (3) We then validate the factors identified from the existing literature by conducting interviews with field experts and expand the list. (4) Finally, we integrate additional factors affecting the HMSC by looking at a serious humanitarian medical crisis (Ebola outbreak in 2014), which pushed the humanitarian medical supply chain to its limits.

FACTORS AFFECTING THE HUMANITARIAN MEDICAL SUPPLY CHAIN

Factors affecting various stages of the medical supply chain related to response have been previously studied in literature. For example, Privett & Gonsalvez (2014) present the findings from interviews and surveys that they conduct with global health supply chain professionals to identify pharmaceutical supply chain challenges. Their research concludes that the top ten challenges are: lack of coordination, inventory, order and warehouse management, demand information, human resource dependency, shortage avoidance, expiration, temperature control, and shipment visibility. Whybark’s study (2007) confirms that factors affecting the medical supply chain are the expiry dates of medicines and their uncertain demand. Hoyos, Morales & Akhavan-Tabatabaei (2015) conduct a literature search on operations research models that have been developed to capture uncertainty in case of disasters. Demand uncertainty, medical supply, healthcare capacity and location of medical centers are recognized as important factors affecting the effectiveness of the supply chain. Salmeron & Apte (2010) discuss the importance of medical staff and supplies for the effectiveness of the supply chain following a disaster response.

Tomasini & Van Wassenhove (2004) describe the humanitarian supply management system (SUMA) that the Pan American Health Organization (PAHO) has developed to provide all relevant information on the flow of donations and purchased medical goods into a disaster area. This system is aimed to help control the entire medical supply chain. Coordination between the different organizations, which depends on their missions, and donations arise as important factors that affect the effectiveness of the supply chain. The above are just some examples of the existing related
work. Based on our extensive study of the humanitarian literature on medical supply chains in case of disaster response, we identify a number of factors affecting the supply chain operations and hence the effectiveness of the HMSC (Error! Reference source not found.).

VALIDATION OF THE FACTORS: INTERVIEWS WITH FIELD EXPERTS

Aiming to better understand the characteristics and factors affecting the effectiveness of humanitarian medical supply chains in post-disaster situations, we conduct a set of interviews with field experts and representatives from humanitarian organizations, as well as with some humanitarian aid responders from governmental agencies, that are involved in such operations. Through these interviews with professionals from the humanitarian sector, we validate the factors identified from the existing literature and, by contrasting them with the reality that these practitioners face, get more in-depth perspective on some of them. Subsequently, we expand the list of factors affecting the HMSC to include our new findings. Error! Reference source not found. presents the updated diagram, where with red font we include the additional factors and relationships between different factors.
CONCLUSION

Challenges outside the direct management of the IHO were classified as external challenges. Within the external challenges, laws and regulations are a great concern within the IHO. National regulations of importation of drugs are increasing and pushing the organizations to procure drugs in the local markets. By procuring nationally the shortage of products is more likely to happen. The IHO also have to assure that the national suppliers meet the WHO standards of quality to avoid counterfeit and substandard products. Furthermore, a competition between IHOs occurs when the number of national wholesalers certificated is not able to meet the demand. Moreover, this competition derives in a price increase of the products. Finally, the assistance can be terminated in cases where the organization is unable to cope with the regulations of the affected country.

Demand uncertainty is another external challenge that impacts the assistance. For example, during the Ebola response, the uncertainty in the demand of PPEs affected the supply chain and created a bullwhip effect. The manufacturers of PPEs did not produce the number of items required by the IHO. The subsequent lack of protective gears for the healthcare workers prevented some organizations from opening new Ebola treatment centers.

The weak infrastructure of the affected country exemplifies another external challenge. The IHOs commonly have to come up with creative solutions to bring the relief items to the
beneficiaries. Indeed, during the Ebola response, the items were quickly moved from the airports to five different warehouses built by some organizations in strategic points of the affected country.

Additional to external challenges are the internal challenges. The latter are directly managed by the IHO. Lack of expertise is a clear example. Hiring an expert on X-rays that visits field operations to provide maintenance to the equipment, exemplifies how the IHO are able to manage this challenge. Capacity of fast deployment is also an internal challenge that affects the effectiveness of the medical assistance. The speed to deploy and install medical equipment, medical vehicles, healthcare structures (including electrical and water installations) is fundamental for the response to an emergency. The medical assistance given in the first days after a natural disaster differs from the assistance provided after several weeks have passed. The cold chain exemplifies another internal challenge. The quality of the drugs could be highly impacted in cases were the cold chain requirements are not met, therefore, the temperature plays an important role during the transportation and storage of drugs.

REFERENCES


Cooperative Humanitarian Logistics Models for Highly Resource-Constrained Settings

Lavanya Marla
University of Illinois at Urbana-Champaign
Urbana, IL, USA

Humanitarian logistics service systems form the underpinnings of a safe, healthy and efficient society. In this work, we focus on operational and policy issues related to resource management and operational interventions in extremely resource constrained settings occurring in these systems. Specifically, we focus on ultra-resource constrained scenarios and settings arising in emergency medical systems (EMS) that necessitate multiple, possibly competing, operators and agencies to cooperate in order to ensure service and public health. In this work we design the operational underpinnings of such cooperative-competitive frameworks, which can be enabled by well-designed information systems, and sensing-enabled smart cities that are designed to explicitly take advantage of such systems.

Motivation

EMS systems have been increasingly stressed over time. The US Bureau of Labor Statistics projects increasing public expectations [2] and a “much faster than average” growth rate of 24% in demand for EMTs and paramedics from 2014-2024 [3]. Likewise, the International Council for Emergency Medicine predicts a steep growth globally, including rapidly urbanizing emerging economies (such as India and South Africa) seeking to emulate the US’s “mature” EMS system [4]. Meanwhile, experts have long suggested overhauling types of response, standards and operational metrics and adding a public health focus; arguing that the current growth rate of EMS systems is unsustainable [5], [6], leading to increasing system stress under new demand and regulatory regimes.

With growing excitement around smart cities and the internet of things, EMS systems have the unprecedented opportunity to leverage connectivity, gather large quantities of data, and make life-saving decisions by interfacing with other public safety systems in real-time [1]. We leverage information-sharing technologies to operationalize those emergency events where resources get overwhelmed, as compared to ‘normal’ operations. These can happen in multiple settings, such as: (i) mass casualty events that require the channeling of multiple medical units to the event, stressing other parts of the system and creating high resource constraints; (ii) emerging economies such as India and South Africa, where the number of ambulances available is far smaller than those in developed economies, and multiple operators exist, giving rise to competition-associated opportunity costs for societal welfare.

Currently, in practice, setting (i) is addressed by ad-hoc cooperation, outlined by mutual aid agreements; between agencies operating in neighboring cities within the same county. In setting
(ii), policy makers are interested in exploring the benefits of (partial to full) cooperation between the multiple competing agencies involved (no such cooperation currently exists). Both settings thus involve the exploration of formal mechanisms involving cooperation that can contribute to social welfare, between potentially competing agencies and resources. The current state-of-the-art is based on mutual aid agreements drawn without data-driven evidence; and therefore, no coherent practices or rules-of-thumb that can guide such practice exist. Evidence from national EMS (NEMSIS) data sets indicates that mass casualty events drastically increase during minor weather events such as snow and this increase is particularly pronounced in rural areas. This suggests there is a need for studying cooperation among multiple agencies using data-driven models even for routine emergencies.

**Challenges**

Our work is the first, to our knowledge, to build a formal mathematical framework to model and compute the value of cooperation among multiple EMS agencies in highly resource-constrained settings. Challenges towards achieving these objectives include:

(i) the lack of such a mathematical modeling frameworks capturing cooperation and competition between networks,
(ii) presence of multiple stakeholders (regulatory bodies, agencies and end users), particularly that each EMS agency operates its resources as a network,
(iii) multiple heterogeneous resources to be allocated (EMS personnel, EMS vehicles) that make the resulting models NP-hard, and
(iv) the presence of significant uncertainty that leads to the requirement of cooperation in the first place, as well as influencing network structure and response times.

**Our approach**

We design and study a mathematical modeling framework that captures cooperation among multiple competing agencies in highly resource-constrained settings. The framework includes both static and dynamic components.

(a) The first level is a static approach that studies for cooperation such that technology-enabled information sharing, and resource sharing can be done, to meet the needs of the multiple agencies and end-users when ultra-resource constraints occur. This allows for positioning resources apriori and understanding what kind of information exchanges may provide benefits. Our formulation is based on the game-theoretic approach of *Cooperative-Competitive decomposition* (referred to as CoCo decomposition [7-9]).

(b) Layered on top of the static decisions, is a dynamic and adaptive mechanism by which resources are allocated in real-time to the networks of the multiple agencies. Because emergency allocation is a dynamic process, this will involve both dynamic information-sharing about emergencies, as well as policies for dynamic re-allocation and dynamic dispatch policies to meet extreme resource constraints as they occur. Within this
mechanism we will embed MDP or ADP-based value functions to describe the performance of each individual emergency responder.

Our two-tier approach has the following features:

(i) First, this mechanism will determine the value of sharing of information and resources, in the static and dynamic layers, and determine the payoffs between each pair of agencies involved.

(ii) Second, our approach incorporates data-driven elements that will use historical data for the static layer and streaming data for the dynamic layer.

(iii) Third, our approach will incorporate ways of learning from the data to compute changing features of the network resulting from changing weather/traffic conditions.

We discuss the static setting here in some detail and leave the dynamic setting for the full paper.

(A) Static/strategic Cooperation setting:

Consider a simple case of a strategic cooperation framework between two neighboring ambulance service systems X and Y, each operating a fleet, engaging in mutual aid. These services could be profit-seeking or public services, with appropriately-defined cost functions. Ambulance allocation, the question of placing ambulances in a network under uncertain demand, is a challenging combinatorial problem in the single EMS case, and more so with two EMS agencies. We formulate the static problem of sharing information and/or resources between the two agencies. The two major questions are: (1) finding a fair and efficient solution among the (combinatorially) many individually rational possibilities in a strategic game, and (2) establishing a play protocol under which strategic players may achieve this outcome.

We use the approach of CoCo decomposition proposed by E. Kalai and A. Kalai [7] - [9]. Coco decomposition is well-suited to the case where there is transferable utility between players, and an ability to make binding agreements and payoff transfers. This approach presents, under a revealed payoff assumption, a coco-value (outcome) which benefits the system the most. For a complete information game with payoff matrices (X,Y) for the two players, the coco decomposition is:

\[
(X, Y) = \left( \frac{X+Y}{2}, \frac{X+Y}{2} \right) + \left( \frac{X-Y}{2}, \frac{Y-X}{2} \right)
\]

\[
Coco\text{-}value(X,Y) = \left( \max_{ij} \frac{x_{ij} + y_{ij}}{2}, \frac{x_{ij} + y_{ij}}{2} \right) + \min \max \left( \frac{X-Y}{2}, \frac{Y-X}{2} \right)
\]

The first component of the coco-value is the cooperative component where the two EMS’s operate as a team, pooling resources, or information (or both). This is the highest payoff received if both players share their payoffs proportionally, based on the contribution to the team. We solve this by considering the demand scenarios for mutual aid, and finding the ‘best’ allocation of ambulances that maximizes total utility; through a simulation-based optimization approach developed by the
authors in previous work [10, 11]. We also demonstrate that alternative approaches such as MEXCLP or MDP or the enhanced hypercube model, may also be embedded.

The second component is the competitive component, played fictitiously. While this game need not be a zero-sum game, this term represents the equilibrium strategy of each player maximizing the difference between its payoff and that of its opponent, also referred to as an advantage game. It can be considered as the advantage that the ‘better placed’ EMS has over the weaker EMS (with no information or resources shared). We can bound the competitive case (under all possible ambulance allocations) using the following formulations.

(B) Extension: Dynamic Cooperation setting:

Given the static solutions achieved in (A), we then examine the dynamic case. We include data-driven prediction algorithms for predicting potentially correlated emergencies occurring during settings of extreme weather and mass casualties, as well as the associated resource allocation dynamic decision-making frameworks. We formulate these using an ADP-based approach to capture the dynamically evolving allocations and dispatch policies, associated costs for each EMS responder agency, and compute time-varying coco-allocations and payoffs enabling the evaluation of both social benefits, the contributions and the benefits achieved by each EMS responder, and the associated payoffs between them.

References

Supply Chain Logistics & Methods
TB4: Games and Collaboration
Thursday 1:00 – 2:30 PM
Session Chair: Nicole Adler

1:00 Multi-Round Combinatorial Auctions for Carrier Collaboration
1Margaretha Gansterer*, 1Richard Hartl, 2Martin Savelsbergh
1University of Vienna, 2Georgia Institute of Technology

1:30 Collaborative Vehicle Routing with Excess Vehicle Capacity in Urban Last-Mile Deliveries
1Joydeep Paul*, 1Niels Agatz, 2Remy Spliet, 2Rene De Koster
1Rotterdam School of Management-Erasmus University Rotterdam, 2Erasmus School of Economics-
Erasmus University Rotterdam

2:00 Competition in Congested Service Networks: The Case of Air Traffic Control Provision in Europe
1Nicole Adler*, 2Eran Hanany, 3Stef Proost
1Hebrew University of Jerusalem, 2Tel Aviv University, 3KU Leuven
Multi-round combinatorial auctions for carrier collaboration

Margaretha Gansterer\textsuperscript{a}, Richard F. Hartl\textsuperscript{a}, Martin Savelsbergh\textsuperscript{b}

\textsuperscript{a}University of Vienna, Department of Business Administration, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
\textsuperscript{b}H Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Drive, NW, Atlanta, GA 30332, USA

Abstract

In horizontal collaborations, carriers form coalitions in order to perform parts of their logistics operations jointly. By exchanging transportation requests among each other, they can operate more efficiently and in a more sustainable way. Through the means of combinatorial auctions, carriers can exchange transportation requests without revealing information about their other tours. They submit part of their requests to a common pool. They are then combined to a set of bundles by a central authority and offered back to all participating carriers. From a practical point of view, offering all possible bundles is not manageable, since the number of bundles grows exponentially with the number of requests that are in the pool. Thus, an attractive subset of bundles has to be generated. In this study we investigate how the concept of multi-round auctions can be adopted in order to contribute to the evaluation of bundle attractiveness. By this we can significantly reduce the number of offered bundles while maintaining solution quality. Three different multi-round auction procedures are developed and assessed. We show how information from a single additional bidding phase can be successfully used, such that the number of offered bundles can be reduced by 60\%, while solution quality is maintained. In the proposed preliminary phase, carriers only have to give their bids on the traded single requests. They do not have to reveal any additional information. We show that this procedure even dominates an approach, where carriers disclose aggregate information on their current tours.

1. Introduction

In the highly competitive transportation industry, carriers need to aim for a maximum level of efficiency in order to stay in business. Fierce competition
brings prices down, therefore profit margins of carriers have declined to an extremely low level. By increasing efficiency, collaborations also serve ecological goals. Thus, public authorities are encouraging companies to collaborate. They aim at reduced road congestion, noise pollution, and emissions of CO₂ and other harmful substances. To approach the goal of maximized efficiency, carriers can, for instance, participate in collaborative networks and trade their transportation requests among each other. This is commonly done by using auction-based exchange systems. In combinatorial auctions requests are not traded individually but are combined into packages, i.e. bundles. The main reason for this is that a specific bundle of requests might have a different value to the partners than the sum of the individual requests. If a bidding carriers price is accepted, the carrier receives the full bundle. The carrier with a rejected bid does not get any item in the bundle. Combinatorial auctions are shown to be effective mechanisms to allocate transportation requests by, e.g., Ledyard et al. (2002), Ackermann et al. (2011), and Berger and Bierwirth (2010). Combinatorial transportation auctions in horizontal collaborations typically follow a 5-phase procedure (Berger and Bierwirth, 2010):

1. Carriers decide which requests to submit to the auction pool
2. Auctioneer generates bundles of requests and offers them to the carriers
3. Carriers give their bids for the offered bundles
4. Winner Determination Problem: Auctioneer allocates bundles to carriers based on their bids
5. Gained profits are distributed among the carriers

The success of such a collaboration clearly depends on the attractiveness of bundles that are offered to the carriers. Thus, a good selection of traded requests (Phase 1) is crucial in order to yield profitable auctions (cf. Gansterer and Hartl, 2016b). Typically, carriers are not willing to trade all their requests, since there are customers they want to serve with their private fleet. However, based on the requests that have been submitted to the auction pool, the auctioneer has to decide on the bundling of requests such that attractive packages can be offered to the carriers (Phase 2). From a practical point of view, offering all possible bundles is not manageable, since the number of bundles is 2ⁿ − 1, where n is the number of requests that are traded.

A common approach to reduce the complexity of auctions is to incorporate multi-round procedures (Dai et al., 2014). In this study we investigate the potential of multi-round combinatorial auctions in less than truckload horizontal collaborations. We develop three different multi-round auction procedures. Our computational study shows that one of these procedures, where only a single additional bidding phase is needed, is clearly dominating in terms of solution quality. We show that the number of offered bundles can be reduced by 60%, while solution quality is maintained. Carriers do not have to reveal any additional information, but information gained in the additional bidding phase can be used to more efficient in generating attractive bundles. This is a significant reduction of the computational complexity, which clearly facilitates the application of combinatorial auctions in real-world collaborations.
2. Multi-round auction procedures

Multi-round auctions generally are intended to offer subsets of the traded items in multiple rounds. Previously gained information can be used to compose the setting for the next round. By this, the bidders are never faced with the full complexity of the auction pool.

We develop three different multi-round procedures for combinatorial transportation auctions:

1. 2-round procedure (2-rA): in the first round carriers bid on single requests only.
2. 2-round procedure (2-rB): in the first round carriers give aggregate information on their current tours. This is done using a grid, which is overlaying the area where requests occur. Carriers have to reveal in which squares of the grid their customers are located. This information is influencing their bid prices.
3. 3-round procedure (3-r): in the first round carriers bid on single requests. In the second round carriers bid on bundles. In the third round, new bundles are composed based on previous bids.

We assume carriers to serve less than truckload paired pickup and delivery requests, which means that each request is associated with a given origin and destination. A carrier starts at the depot, visits a given set of pickup and delivery nodes and returns to the depot again. The objective is to minimize total travel time. Carriers’ tours cannot exceed a given carrier specific tourlength. In the bundling phase, the auctioneer has to compose packages of requests that are of high attractiveness to carriers. In Gansterer and Hartl (2016a) it has been shown that the attractiveness of bundles can be approximated by geographical information, which includes the density, the isolation, and the total travel time needed to visit all requests in a bundle.

However, for the multi-round approach we develop several methods for building bundles based on information gained in previous bidding rounds. We show that the best solutions are found, if we consider the following term in the bundle evaluation procedure:

\[
\max_{c \in C} \sum_{r \in R} p_{cr} \frac{\max_{c \in C} \sum_{r \in R} p_{cr}}{\sum_{c \in C} \sum_{r \in R} \max(p_{cr})},
\]

where \(C\) and \(R\) are the sets of carriers and requests, respectively, while \(p_{cr}\) is the bid price of carrier \(c\) on single request \(r\). By this formula we relate the bids of a carrier to the bids of his collaborators. A bundle gets the maximum evaluation of 1, if it only contains requests, where carrier \(c\) submitted the highest bids. This bundle evaluation is combined with the geographical assessment mentioned above.

In the 3-round procedure, the first auction round is identical to that of the 2-round procedures. In the third round we generate a new set of bundles, for which we again investigate several methods. We, for instance, calculate synergy...
Table 1: Percentage deviation in collaboration improvement compared to the single-round auction. A positive value implies that the collaboration improvement found with the new auction procedure was higher than that of the single-round auction. In the second column we give the number of bundles that were offered to the carriers.

<table>
<thead>
<tr>
<th></th>
<th>#Bundles</th>
<th>O1</th>
<th>O2</th>
<th>O3</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-round</td>
<td>500</td>
<td>47.3</td>
<td>69.2</td>
<td>65.3</td>
<td></td>
</tr>
<tr>
<td>2-rA</td>
<td>200</td>
<td>-2.09%</td>
<td>-1.95%</td>
<td>0.77%</td>
<td>0.30%</td>
</tr>
<tr>
<td>2-rB</td>
<td>200</td>
<td>-1.84%</td>
<td>-3.38%</td>
<td>-17.99%</td>
<td>-6.51%</td>
</tr>
<tr>
<td>3-r</td>
<td>200</td>
<td>-7.50%</td>
<td>-8.77%</td>
<td>-1.50%</td>
<td>-5.93%</td>
</tr>
</tbody>
</table>

values for each request pair in the auction pool. If a request pair is included in many highly ranked bundles, this request pair gets a high synergy value. We then compose bundles such that good request pairs are offered in the same bundle.

3. Computational study

We consider different scenarios in terms of (i) degree of competition and (ii) distance of pickup and delivery points to the carriers’ depots. We assume that 3 carriers operate in overlapping but not identical customer regions. There are 3 types of instances depending on the degree of customer area overlaps. For each instance, we generate equidistant carrier depots with a distance of 200. Pickup and delivery points are randomly generated within a radius of 150 (O1), 200 (O2), and 300 (O3). Each carrier initially holds 15 requests. We generate 20 instances for each scenario.

In Table 1 we compare the results of each multi-round auction to the results of the single-round auction, where only one bidding phase is conducted.

The results show that the 2-round procedure, where 200 bundles are offered, yields on average an even better solution than the single-round auction, where the number of offered bundles is 500. It should be noted that the only difference between these auction procedures is a preliminary bidding phase where carriers have to give their prices for single requests. The alternative 2-round procedure is dominated. It does not yield the same solution quality, although carriers are revealing information on their current tours. Also the 3-round procedure is clearly dominated.

4. Conclusion

In this study, we investigated the potential of multi-round combinatorial transportation auctions in horizontal collaborations. Since the number of possi-
ble bundles increases exponentially in the number of traded requests, real-world settings can only be run, if the auctioneer is able to offer a subset of possibly attractive bundles. However, since carriers do not want to reveal sensitive information, the auctioneer has to compose these subsets having only incomplete information. We show how information from a preliminary bidding phase can be successfully incorporated, such that the auction’s computational complexity can be reduced significantly. In this preliminary phase carriers only have to give their bids on the single requests. They do not have to reveal any additional information. We show that this procedure dominates an alternative approach, where carriers have to disclose some information on their current tours. Also a 3-round procedure is dominated. However, using the proposed 2-round procedure, we can reduce the number of offered bundles by 60%, while solution quality is maintained. This makes combinatorial auctions a powerful tool for real-world horizontal collaborations.

References


Collaborative vehicle routing with excess vehicle capacity in urban last-mile deliveries

Joydeep Paul\textsuperscript{a}, Niels Agatz\textsuperscript{a}, Remy Spliet\textsuperscript{b}, René de Koster\textsuperscript{a}

\textsuperscript{a}Rotterdam School of Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062PA Rotterdam, The Netherlands

\textsuperscript{b}Erasmus School of Economics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062PA Rotterdam, The Netherlands

1 Introduction

Traffic congestion has become a serious issue in many urban areas around the world. According to Cebr [1], total costs for congestion in the UK, France, Germany, and USA are forecasted to rise from $200 billion in 2013 to $293 billion by 2030 - an alarming 46% increase. At the same time, it has been observed that vehicles are running with a lot of empty space. The load factor for trucks in Europe is around 50% [2] and a similar trend is seen in other parts of the world as well. Consolidation of different transportation flows is one way to increase the load factor of the vehicles in the logistics network. It is, however, challenging to consolidate flows between different business entities that have their own transportation planning. Feedback from industry suggests that the cost of collaboration (infrastructure development to align systems) often outweighs the advantages. As the logistics landscape is becoming more dynamic, companies do not want to commit to long-term collaborations. Fortunately, recent technological advances allow more dynamic forms of collaboration by facilitating easy exchange of information.

We consider a dynamic collaboration between two logistics service providers, where one (focal carrier) has the possibility to piggyback on the routes of the other service provider (external carrier) by utilizing the unused capacity. Our research is motivated by a collaboration with one of the leading grocery retailers in the Netherlands. The grocery retailer has physical stores which also act as pick-up points for items ordered online. Each of the channels (replenishment and online fulfillment) has its own warehouse and independent route planning. As a result, stores are being visited by vehicles for replenishment of inventory and also, by vehicles to supply the pick-up points. In this setting, the replenishment channel acts as the external carrier which fixes the route plans in advance and communicates the excess capacity in each route to the online channel which acts as the focal carrier. Since there is overlap in the customer locations, it provides an opportunity for the focal carrier to redirect the demand of the common customer locations to the external carrier route, thereby reducing the total vehicle-miles of the system and also the number of visits to the customer locations. However, there will be additional costs of moving items to the external carrier warehouse and handling costs to load the items into the vehicles of the external carrier route.
It is also observed that the excess capacity in the external carrier routes varies from day to day, hence it is not always cost-feasible to consolidate demands of customer locations.

In this paper, we focus on the focal carrier’s decisions to assign customer locations to the external carrier and construct routes for the remaining customer locations. This problem can be characterized as a selective vehicle routing problem as we do not have to visit all locations. In the literature [3] [4], the selective vehicle routing problem has an exogenous revenue associated with each customer locations, and objective is to maximize the revenue by visiting a subset of customer locations in a given time. In our problem, there is no explicit exogenous revenue parameter but the selection of customer locations, not to be visited, is constrained by the excess capacity of the external carrier routes and inter-warehouse transfer cost. Our contribution is threefold. First, we introduce a new dynamic consolidation concept in the domain of consolidation strategies among logistics partners. Secondly, we develop different heuristics to solve large instances. Finally, we present a numerical study to investigate the benefits of such consolidation strategy in different settings.

2 Problem Definition

We model our selective routing problem on a graph where the node set $N$ represent the customer locations of the focal carrier and every route is a Hamiltoninan cycle starting and ending at the warehouse. Each vehicle has capacity $Q$ and corresponds to a route. Each node $i \in N$ has a demand $q_i \leq Q$. Each node can be served by one of its own vehicles or redirected (if feasible) to one of the external carrier routes. The handling costs (time) for serving a customer location are the same for each vehicle in the focal and external carrier. We do not allow partial redirections of load from the focal to the external carrier.

Let $r$ be a route of the external carrier and let $N(r)$ be the set of nodes served by that route. The unused capacity of a vehicle in an external route $r$ is referred to as its excess capacity and denoted by $e_r$. For each external route $r$, a set $S_r \subseteq N(r)$ is called flexible node set if it is feasible to be redirected from the focal carrier route, i.e. if $\sum_{i \in S_r} q_i \leq e_r$.

We define a feasible load assignment as the assignment of nodes to the external carrier so that the available capacity of the external carrier’s routes is respected. So, a feasible load assignment involves selecting a flexible node set $S_r$ for every external route $r$.

If a node is served by the external carrier instead of the focal carrier, we need to employ an inter-warehouse vehicle to transfer the items from the focal to the external carrier’s warehouse. There is a fixed cost $f_w$ incurred per inter-warehouse trip with vehicle $z$ of capacity $Q_z$. It is possible to split customer loads on inter-warehouse trips. This means that if $S$ is the set of nodes redirected to the external carrier, then the required number of inter-warehouse trips is given by $\lceil \sum_{i \in S} q_i / Q_z \rceil$.

The objective is to determine the optimal load assignment and corresponding routing so as to minimize total vehicle-miles of the focal carrier routes and the inter-warehouse trips. For the consolidation to be beneficial, the inter-warehouse cost must be less than the savings attained by redirecting the nodes’ demand to the external carrier routes.
3 Methodology

3.1 MIP Solver - Mathematical Modeling

We modify the two-index formulation for capacitated Vehicle Routing Problem (VRP) as given by Toth and Vigo [5] to include the additional constraints on excess capacity and inter-warehouse movement. However, since the VRP is a difficult problem, only small instances of our problem can be solved with any standard MIP Solver.

3.2 Two Stage Solution Approach

We decompose the problem into a selection phase to determine feasible load assignment(s) and routing phase to evaluate the costs of these assignment(s).

3.2.1 Determination of feasible load assignment

Exact:
We enumerate all feasible load assignments and evaluate each to find the optimum. As mentioned in Section 2, a feasible load assignment involves selecting a flexible node set $S_r$ from each external route $r$. Enumerating all feasible load assignments requires evaluating all possible subsets of $N(r)$ for every external route $r$. So, the enumeration grows exponentially with the number of nodes in each route and the number of external routes, which makes enumeration of all options only possible for small instances. Therefore, we introduce two heuristic methods to find an efficient and feasible load assignment.

Heuristic:

Greedy Removal: We solve a VRP (using an ALNS heuristic [6]) to form initial routes for the focal carrier route which visit all the nodes. We calculate the savings (the cost of edges connecting the node) in the routing costs associated with each feasible node redirection. Then, we choose the one with the highest savings to redirect to the external carrier. We repeat the process of calculating the savings and the selection, until no more nodes can be redirected.

Knapsack: We can approximate the savings of redirecting a node to the external carrier in different ways. One simple proxy is the distance of the node from the depot. The further it is from the depot, the more advantageous it is to redirect. Based on this approximation, the nodes are selected to maximize the savings of load assignment while respecting the available capacity of the external carrier routes. This gives rise to a problem that has a similar structure as the knapsack problem, which can be solved very fast, even for large instances.

3.2.2 Evaluation of a feasible load assignment

For a given load assignment, we know the nodes that need to be served by the focal carrier. The cost of the solution is the sum of the routing cost and the inter-warehouse movement cost. Routing cost can be determined by using an exact or heuristic method to solve the VRP.
• **Exact:** For small instances, we solve the capacitated VRP for the focal carrier route using MIP Solvers like Gurobi by introducing simple capacity cuts iteratively.

• **Heuristic:** We solve the VRP for the selected nodes of the focal carrier route using the ALNS algorithm [6].

### 3.3 Local Search

We implement a local search heuristic to solve the problem. The heuristic has three different phases. In the *initialization* phase, we select a feasible load assignment from each cluster randomly and then use ALNS [6] to generate routes corresponding to the selected nodes of the focal carrier. In the *intensification* phase, we use simple operators like move and swap to improve the solution locally. Finally in the *diversification* phase, we destroy parts of the solution and repair the same to obtain new and better solutions.

### 4 Experiments and Results

**Experiment Setting** We generated a set of test cases using the VRP instances from VRP-lib [7]. The best known solutions for these instances are known and are used as reference for calculating the savings achieved by consolidation. To evaluate our heuristics, we use the VRP-lib instance of size 32 nodes (including the depot). The routes of the external carrier are generated by solving a VRP with the ALNS heuristic. The excess capacity $e_r$ of each route $r$ is generated uniformly between 0 and 0.25 times of total focal demand of the nodes in that route, which means, on average, an external carrier route of 4 nodes can accommodate demand of 0 to 1 node of the focal carrier. We generate 10 realizations of the external courier routes and the results reported in the next section are average values of these 10 realizations.

**Results** Table 1 shows the performance of the different heuristics (as described in Section 3.2 and 3.3) with respect to the average gap from the optimal solution ($\Delta$), the number of times the optimal solutions are found out of 10 realizations (# OPT) and the largest optimality gap (max. $\Delta$ (%)). We also record the number of cases in which the consolidation strategy leads to positive savings (# SAV > 0), that is, in how many cases at least, some consolidation through redirection was achieved. This provides us insight into the dynamic nature of the problem. Based on the excess capacity and the routing of the external carrier routes, the ad-hoc decision to redirect or not has to be made. In Table 2, we show the number of nodes (# Nodes) whose demand was redirected to the external routes and the number of inter-warehouse vehicles (# InterVeh) needed, based on the redirection strategy of the three heuristics. We observe a decrease in the number of vehicles (# Veh) and distance (Distance %) traversed in the focal routes in the consolidated network. Although the Knapsack heuristic redirects more number of nodes to the external routes as compared to the Local Search heuristic, the total distance reduction is more for the Local Search heuristic as the distance reduction depends on the selection of nodes to be redirected. The Local Search heuristic performs better than the other two heuristics in solution quality.
Table 1: Benchmarking

<table>
<thead>
<tr>
<th></th>
<th>$\Delta$ (%)</th>
<th># OPT</th>
<th>max. $\Delta$ (%)</th>
<th># SAV $&gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greedy + Exact</td>
<td>2.16</td>
<td>1/10</td>
<td>3.8</td>
<td>5/10</td>
</tr>
<tr>
<td>Greedy + Heuristic</td>
<td>2.16</td>
<td>1/10</td>
<td>3.8</td>
<td>5/10</td>
</tr>
<tr>
<td>Knapsack + Exact</td>
<td>1.16</td>
<td>1/10</td>
<td>3.7</td>
<td>8/10</td>
</tr>
<tr>
<td>Knapsack + Heuristic</td>
<td>1.16</td>
<td>1/10</td>
<td>3.7</td>
<td>8/10</td>
</tr>
<tr>
<td>Local Search</td>
<td>0.01</td>
<td>9/10</td>
<td>0.1</td>
<td>9/10</td>
</tr>
<tr>
<td>No - Collaboration</td>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Consolidated Network

<table>
<thead>
<tr>
<th></th>
<th>To external</th>
<th>$\Delta$ focal routes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Nodes</td>
<td># InterVeh</td>
</tr>
<tr>
<td>Greedy + Exact</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Greedy + Heuristic</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Knapsack + Exact</td>
<td>9.5</td>
<td>1</td>
</tr>
<tr>
<td>Knapsack + Heuristic</td>
<td>9.5</td>
<td>1</td>
</tr>
<tr>
<td>Local Search</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

5 Conclusion

We have introduced a new strategy for collaborating with logistics partners on an ad-hoc basis utilizing the excess capacity of vehicles. The initial results from the experiments look promising and suggest that savings can be achieved by dynamic collaboration between two logistics partners using the excess capacity of one into another. We plan to run similar experiments on bigger instances to show the savings that can be achieved under different scenarios. We have considered the collaboration scenario among two logistics players in this work, but the same framework can be extended with multiple players where we will observe multiple external-focal carrier collaborations.

References

Service providers face the continuous challenge of adapting capacities to the fluctuations in demand patterns of congestion sensitive customers. Air traffic demand has been estimated to increase continuously over the next two decades which requires the adoption of technology to navigate aircraft accurately and safely through the skies otherwise the additional demand will not be served without substantial delays. An important aspect of air traffic control (ATC) draws from the network dimension whereby, in Europe, each section of the network is managed by a single service provider. However, flights may cross multiple sections of the airspace, hence need to be served by several providers seamlessly. In addition, given the multiple potential routes from origin to destination, it is clear that European ATC service providers are also competing in certain markets. Consequently, ATC providers are simultaneously required to cooperate to ensure seamless service and to compete for certain traffic routes.

The ATC market is one example in which service providers compete over a geographically congested network. The ATC provider is both a monopolist that could maximize profits by setting capacities and prices for use of their own airspace and a competitor with their neighbors for transit traffic. The setting becomes more complicated because the airlines are non-atoministic in that each has market power by generating a non-negligible fraction of the total demand. Airlines maximize their profits by choosing the cheapest route to fly whilst internalizing their own congestion costs. The management issues caused by this complex market include high costs due to fragmentation of ATC service provision, slow technology adoption, lack of standardization in ATC services and inefficient scale of operation. Consequently, a regulatory policy may involve price caps, investments in technology R&D and the promotion of horizontal integration for the purpose of achieving cost efficiency simultaneously.

More generally, competition between service providers may be modelled using a directed flow network in which each arc represents service provided by a specific supplier and each customer demands path flows connecting specified origin-destination pairs. In this research we model such settings within a two-stage network congestion game: in the first stage each service provider sets charges on the services (arcs) they provide to maximize profits; in the second stage each customer chooses the desired flows to minimize the sum of service charges and congestion dependent operating costs, including the possibility of partial flows or not using any service when the associated costs are too high. The modeling approach is the first research known to the authors that considers general networks, demand for multiple origin-destination pairs and oligopolistic markets in both stages of the game. It is also the first to model non-atoministic airlines with market power in the second stage. We use this model to analyze whether competition between service providers may lead to efficiency in the sense of minimizing total social costs and/or lower service charges for the benefit of the airlines. The application of the congestion game with pricing thus enables an evaluation of multiple market design scenarios including the impact of deregulation, incentive based price caps, different forms of co-operation between players and the introduction of new technology.

One of the aims of this research is to search for paths to accelerate change in the European air traffic control (ATC) sector, in line with the Single European Skies initiative and the SESAR program. Slow adoption of technology occurs in many industries and the specific reasons in the ATC industry in Europe include: the fragmentation of the ATC providers, the home-bias of each country for the national provider, the monopolistic nature of some of the ATC services, the network component of most ATC services and the split incentives which require the ATC providers to invest in new technology without enjoying any direct benefits beyond potentially increased demand. The fact that the ATC providers are required to bear investment costs and to invest effort through increased coordination, while the airlines are the main beneficiaries, will most likely delay the implementation of new technology as advocated by the European Commission. The reality is that both the United States, through the NextGen
program, and the Far East, through the CARAT program, are dealing with similar issues in terms of technology adoption or lack thereof.

I. MODELING APPROACH

We develop a network congestion game to test a series of scenarios in order to analyze potential paths for change in air traffic management in Europe. We assume a two stage game in which the ATC providers make decisions in the first stage and then airlines respond in a second stage. The first stage requires the ATC providers to set their user charges according to specific objectives, such as revenue or profit maximization. In the second stage of the game, the airlines choose their flight paths such that they minimize their operational costs. The costs include variable costs, covering labour and fuel, congestion costs and the ATC charges, all of which are impacted to some degree by the ATC provision. For example, the more direct the flight path, the lower the fuel and staff costs for the airline. Equivalently, the lower the congestion in airspace and the higher the capacity, the lower the congestion costs for the airlines, which contribute about 10% to the total airline operating costs in Europe. The direct ATC user charges add an additional 5 to 8% to the airlines’ operating costs. Consequently, the decisions of the ATC providers will impact the airlines directly and these are the two main players in the game.

The two stage game models en-route and terminal air traffic control providers that set peak and off-peak charges and in the second stage airlines that choose flight paths given an airline schedule and the charges from the first stage. A quadratic objective function (due to congestion costs) with linear constraints describes the airline level decisions and the ATC providers simultaneously choose their charges as a best response to the choices of their competitors, taking as given the optimal airline flows, thus leading to a Nash equilibrium. The scenarios analyzed in the model include (i) the impact of privatization and deregulation; (ii) defragmentation of the set of current providers; (iii) introduction of technology via the common projects and SESAR step 1; and (iv) the regional forerunner approach in which ATC providers and a specific airline co-operate.

In an initial stage, the regulators set the rules with respect to institutional form, types of co-operation permitted, price regulation and levels of technology implemented, creating a series of potential market design scenarios. These alternative scenarios are compared with respect to service charges and according to a total social cost function applied to the sub-game perfect equilibrium outcomes of each scenario analyzed.

The network underlying the congestion game is composed of a set of origin, transit and destination nodes, and a set of arcs representing services offered. We use the following network definitions:

\[ P \] finite set of origin/destination nodes with indices \( o, d \)
\[ T \] finite set of transit nodes
\[ N \subseteq N \times N \] set of all nodes, \( N = P \cup T \), with indices \( i, j \)
\[ A_s \subseteq N \times N \] set of arcs belonging to service provider \( s \)
\[ A \] set of all arcs, \( A = \bigcup_s A_s \), with index \( a = (i, j) \)
\[ \delta_a \] weight (length/time) of service in arc \( a \)
\[ W = \{1,2\} \] set of peak and off-peak time windows, with index \( w \)

for the service providers and airlines we use the following definitions:
\[ L \] finite set of airlines, with index \( l \)
\[ S \] finite set of service providers with index \( s \)
\[ s(a) \] service provider controlling arc \( a \)
\[ K_{law} \] maximal flow for customer \( l \) in arc \( a \) at time window \( w \)
\[ D_{lod} \] demand of customer \( l \) for service from origin \( o \) to destination \( d \)

and we use the following definitions of costs and charges:
\[ c^0 \] customer \( l \)'s operating cost per weight unit in arc \( a \)

\[ c^R \] customer \( l \)'s cost per weight unit from off-peak service in arc \( a \) at time window \( w \)

\[ c^G \] customer \( l \)'s congestion cost per weight unit in arc of service provider \( s \)

\[ c^E \] airlines' outside option cost to service from origin \( o \) to destination \( d \)

\[ c^S \] service provider \( s \)'s variable cost per weight unit in arc \( a \)

\[ t^w \] service provider \( s \)'s price cap per weight unit at time window \( w \)

Finally, we use the following sets of decision variables:

\[ \tau^w \] service provider \( s \)'s charge per weight unit in arc \( a \) at time window \( w \)

\[ f^*_{loda} \] customer \( l \)'s service flow in arc \( a \) for the purpose of origin-destination pair \((o,d)\) at time window \( w \)

\[ f^T \] customer \( l \)'s non-flow from origin \( o \) to destination \( d \)

Prior to the game, the regulator may set price caps, \( \tau_{aw}^0 \), for service providers, enforce legislation with respect to horizontal integration across service providers and set levels of technology required to be implemented by the actors. In stage one, service providers set their charges, \( \tau_{aw} \). We model the service providers as profit maximizers, with or without a price-cap regulation. Each service provider best responds to the choices of its competitors, taking as given the equilibrium service flows \( f^*_{loda} \) that will be chosen in the second stage of the game, thus leading to a sub-game perfect Nash equilibrium. We define a profit maximization objective function per service provider \( s \). Model (1) includes a set of constraints in which the charges are price capped, to be included where relevant.

\[
\text{Max}_{\tau_{aw}} \sum_{a \in A_s} \sum_w (\tau_{aw} - c^S_{wa}) \beta_a \sum_{od} f^*_{loda} \\
\text{s.t.} \quad \tau_{aw} \leq \tau_{aw}^0 \forall a \in A_s, w \in W
\]

The customer cost functions, which are modelled in the second stage of the game with linear latency costs, are composed of five categories, all of which are impacted to some degree by the service providers. This quadratic objective function, equation (2), includes operating costs \( c^0_{la} \), cost \( c^R_{lawa} \) from flying off-peak (or loss of consumer surplus for the airlines), a congestion cost \( c^G_{ls} \) and service provider charges \( \tau_{lawa} \). Additionally, in order to account for elastic demand, there exists an outside option flow, \( f^T_{loda} \), which represents the choice to reduce service, with cost \( c^E_{od} \perunit \) per flow unit, which will be preferred if the total costs of being served are too high.

\[
\Psi_l \equiv \sum_w \sum_{a \in A} \left[ c^0_{la} + c^R_{lawa} + c^G_{ls(a)} \sum_{l'da} f^*_{l'da} + \tau_{ls(a)aw} \right] \beta_a \sum_{od} f^*_{loda} \\
+ \sum_{od} c^E_{od} f^T_{loda}
\]

Two alternative solutions are modeled for the second stage: either a system optimal outcome as described in equations (3) to (6) or a user equilibrium outcome in which the objective function is replaced by equation (3'). In the system optimal approach, a central planner chooses the service paths and timing (peak or off-peak) for all airlines simultaneously to minimize total customer costs. This solution would minimize total social costs were the service charges to equal the marginal service cost. The quadratic objective function (3) thus minimizes the costs of all airlines, taking into account total operating costs and revenue losses from being served in the off-peak.

\[
\text{Min}_{f_{loda}, f^T_{loda}} \sum_l \Psi_l \\
\text{s.t.} \\
\sum_w \sum_{j \in A} f_{loda(j)} - \sum_{j \in A} f_{loda(j)} = D_{loda}, \forall l \in L, \forall o, d \\
\sum_w \sum_{j \in A} f_{loda(j)} = D_{loda}, \forall l \in L, \forall o, d
\]
\[
\sum_{j \in (i,j) \in A} f_{lod(j,i)}w - \sum_{j \in (i,j) \in A} f_{lod(i,j)}w = 0, \forall l \in L, w \in W, o, d, i \in N (i \neq o, d)
\]

\[
\sum_{o,d} f_{lodaw} \leq K_{law}, \quad \forall l \in L, \forall a \in A, \forall w \in W
\]

\[
f_{lodaw} \geq 0, f_{lod}^T \geq 0, \forall l \in L, o, d \in N, a \in A, w \in W.
\]

Constraint (4) sums the incoming less the outgoing flows to be equal to the (negative) demand at the (origin) destination and zero when using a transit point. The total flows are reduced by those that have been dropped via the outside option. Constraint (5) restricts the number of flows in the peak hours per customer. Constraint (6) ensures non-negativity of the flows and non-flow.

In a user equilibrium outcome, we assume that each customer chooses paths and time windows taking into account its own costs only and taking the flows of the other airlines as given. Specifically, each customer \( l \) considers only its own congestion costs and ignores the external congestion costs imposed on the other airlines. Hence the flows may be less balanced than those of the system optimal approach. This solution would minimize total social costs were the service charges to equal the marginal service cost plus the external congestion costs. A user equilibrium is evaluated under the same set of constraints, (4) to (6), but the objective function is adapted as shown in (3').

\[
\text{Min}_{f_{lodaw}, f_{lod}} \Psi_l
\]

Finally, we compare the social costs to be minimized across all scenarios in order to search for the most appropriate equilibria outcomes considering both sets of actors. The social cost function sums all customer costs less the service provider profits.

II. RESULTS

We apply the model to a case study as depicted in Figure 1. We analyse six en-route air traffic control providers and nine terminal providers based in Western Europe which covers 50% of the traffic served across European airspace. In the second stage, we model five airline carriers, including three alliances, one low cost carrier and one representative non-European carrier. In order to shed light on the potential impact of changes in institutional or regulatory arrangements we study five groups of scenarios. The first group is the base-run scenario in which we reproduce the 2011 equilibria outcome. In scenario group 2, we analyze the likely outcome were a central manager to provide all ATC services, in a similar style to that of the U.S. federal aviation system. In scenario group 3, we highlight the potential impact of the functional airspace blocks which are the equivalent of horizontal integration across groups of ATC providers. We assume that there will be no savings in labor costs or reduction in air control centers due to the power of the labor unions and the politics of sovereign protection but savings of up to 30% are possible in the sum of the fixed costs due to joint purchasing power and standardization of processes. In addition, we test the assumption that variable cost savings are possible due to a one third reduction in the joint support staff in addition to the joint procurement cost savings. In scenario group 4 we analyze the potential impact of technology on the equilibrium outcomes by modelling the expected costs and benefits of new technologies to both the ATC providers and the airlines. In scenario group 5, under vertical integration, an ATC provider and its relevant hub airline are assumed to adopt new technology and via the best-equipped best-served scheduling rule are able to achieve the benefits of the technology locally. Within each group of scenarios, we analyse four sub-cases including the user equilibrium cost recovery constraint, system optimal with cost recovery constraint, user equilibrium price-cap approach and the profit maximization solution. We use dynamic programming algorithms to solve the relevant model per ATC provider and airline, continuing until no actor changes the values of the decision variables. The second stage quadratic, convex,
A continuous problem is solved iteratively using CPLEX version 12-6-2 and the first stage is solved iteratively given the second stage results using line search methods. Finally, we note that all scenarios are analyzed using 2011 demand and subsequently 2020 and 2030 Eurocontrol forecasted demand estimates which suggest that demand is expected to increase by 19.5% and 38.7% respectively as compared to 2011.

![European air traffic control network case study](image)

**Figure 1: European air traffic control network case study**

We identify a lack of incentives to encourage efficiency at the Member State level, which is overcome when intra-European traffic is analysed at the federal level. Consequently, centralized services are likely to lead to more direct flight paths for the airlines and to economies of scale at the ATC provider level, which may reduce costs. We learn that there is insufficient competition across flight paths in different regions to permit the removal of economic regulation. Competition is only likely to arise when providers are in a position to compete for services over the same set of flight paths, for example through virtual centres or through trajectory-based provision. Second, horizontal integration via functional airspace blocks is unlikely to facilitate cooperation across providers due to the lack of financial incentives. If functional airspace blocks are required to set a single price across their entire network, an average price is likely to lead to some airlines winning and other losing whereas prices set at the lowest current ATC provider level lead to lower profits for the functional airspace blocks once the ATC providers combine. Consequently, functional airspace blocks would need to set differential prices across their airspace. Furthermore, the cost of standardizing equipment in the shorter term will likely require subsidies or higher prices, which is in direct opposition to current European policies. Third, in order to encourage technology adoption that will reduce airline costs according to SESAR, the ATC providers ought to be permitted to increase their charges in the range of 10 to 20%. Consequently, the current system of incentives needs to be altered in order to accelerate change and encourage the adoption of SESAR technologies. Fourth, the regional forerunner approach akin to vertical integration may lead to piecemeal adoption of new technologies as in specific cases, this would be to the benefit of both the ATC provider and the local airline, for example Lufthansa and DFS. In other words, a more piecemeal adoption of technology where shown to be relevant may lead to greater change than the current top-down legal approach taken to date. Institutionally, a clear separation of the ATC providers from the Member States and subsequent franchising of the support services and ATC services could further encourage efficiency, consolidation and technology adoption.
Supply Chain Logistics & Methods
TC4: Delivery Service Network
Thursday 2:45 – 4:15 PM
Session Chair: Niels Agatz

2:45  An On-Demand Same-Day Delivery Service Using Direct Peer-to-Peer Transshipment Strategies
1Wei Zhou, 2Jane Lin*
1CH Robinson, 2University of Illinois at Chicago

3:15  Balancing Availability and Profitability in E-Fulfillment with Revenue Management and Predictive Routing
1Catherine Cleophas*, 2Jan Fabian Ehmke, 2Charlotte Köhler, 1Magdalena Lang
1RWTH Aachen University, 2European University Viadrina

3:45  Heuristic Approaches to the Same-Day Delivery Problem
1Alp Arslan*, 2Niels Agatz, 2Rob Zuidwijk
1Rotterdam School of Management, 2Rotterdam School of Management Erasmus University
An On-demand Same-Day Delivery Service Using Direct Peer-to-Peer Transshipment Strategies

Wei Zhou\textsuperscript{a}, Jane Lin\textsuperscript{1, b}

\textsuperscript{a}C.H. Robinson, Inc., Eden Prairie, MN, U.S.A.
\textsuperscript{b}Department of Civil and Materials Engineering & Institute for Environmental Science and Policy, University of Illinois at Chicago, U.S.A.

1. Introduction

The fast growing e-commerce industry is rapidly changing the landscape of the retail business, a 4.5 trillion dollar industry for which on-line sales could account more than 10\% by 2017 (U.S. Census Bureau report, 2015). This has resulted in increasing volume of goods transportation as well as increasing demand for fast, cheap and reliable delivery service by consumers. On the other hand, the rapid advances in wireless communication and ubiquitous mobile computing have provided real opportunities for revolutionizing the traditional transportation market by enabling on-demand mobility services fast and cheaply through improved matching of transportation demand and services (supply) and fuller utilization of vehicle capacity.

In particular, as the same-day delivery service becomes increasingly popular among e-commerce customers, it demands fast and cheap delivery service. Moreover, the same-day delivery requests may appear throughout the day, further demanding flexible on-demand service. Thus emerges an on-demand, same-day (ODSD) delivery service market.

This has placed a great deal of pressure on the traditional hub-and-spoke package delivery paradigm used in the industry including FedEx, UPS, and USPS. The traditional package delivery operation involves first collecting all delivery items from the customers in the service areas pre-assigned to a service fleet and moving them to a sorting facility where items are then sorted before they are loaded up to the service fleet again and delivered. As such, the sorting operation is in effect a bottleneck in transshipment and there is no guarantee for a same-day service in this traditional hub-and-spoke service paradigm. A naive solution to meet the ODSD service demand is direct shipping, in which a vehicle picks up a load and delivers directly to its receiver without stops (Burns et al. 1985). Direct shipping is a most responsive strategy but requires additional vehicles in the fleet to serve the demand. This results in additional capital investment, daily operation, and maintenance of a larger fleet, which means increased cost to the business and customers. It also results in increased number of vehicle trips, which may exacerbate the already congested urban streets and air pollution. More importantly, direct shipping is not well suited with economies of scale.

In light of the emerging ODSD urban delivery service market and challenges described above, this study proposes, formulates, and evaluates a new ODSD delivery strategy using direct peer-to-peer transshipment (P2PT)\textsuperscript{2}. The proposed P2PT involves en-route package relays among multiple couriers to extend beyond the normal service range of a single courier in order to reach the package's final destination; this is done directly among the couriers without going through a major transshipment facility such as a sorting center, a distribution center, or an urban consolidation center (UCC). Hence it is a peer-to-peer transshipment. Without having to route through a transshipment facility, P2PT could remove the transshipment bottleneck and warrant an ODSD delivery service after proper scheduling of the visiting orders and routing of the vehicles; it is a sharing economy by taking advantage of under-utilized vehicle capacity already in the service fleet to accommodate the ODSD delivery packages, and thus may gain economies of scale than direct shipping in the ODSD delivery service.

\textsuperscript{1}Corresponding author. Email addresses: zhou.wei.iris@gmail.com (W. Zhou), janelin@uic.edu (J. Lin)
\textsuperscript{2}It is assumed in this study that these on-demand same-day delivery packages are small in size and light in weight.
P2PT can be realized via effective collaboration and coordination among couriers. This is illustrated graphically in Figure 1. S1 and S2 are two neighboring service zones each serviced by a courier. Suppose an ODSD service request arrives at some point of the day consisting of a pickup location P within the service range of S1 and a delivery destination D within the service range of S2. In P2PT, the package is picked up at P by S1 and transferred to a relay point T with a drop box accessible by both S1 and S2. The package is then stored in the drop box at T and later picked up by S2 and delivered to its destination D in S2. If this transshipment involves more than one relay, then S2 carries the package to the next relay point to be picked up by the next courier, and so on and so forth till the package reaches its final destination. In this study, each courier has a set of candidate relay points to choose from and the selection depends on who is the immediate downstream courier to relay to.

![Figure 1. Graphic illustration of peer-to-peer transshipment in P2PT](image)

Formation of the partnership between two neighboring couriers can be pre-determined or ad hoc. In the former, the partnership is fixed. In the latter, the partnership is formed temporally as needed to best carry out the transshipment; once the job is done, the partnership dissolves. In this study, the ad hoc relay partnership is considered and formulated. The underlying principle of the ad hoc P2PT is better utilizing transportation resources (e.g., vehicle capacity) and sharing cost through sharing economy.

There are several dimensions of coordination that take place in P2PT: (a) formation of the ad hoc partnership - who to team up between each relay along the way, (b) joint selection of relay points among all the couriers involved, and (c) coordination of package drop-off and pickup at a relay point. In this paper all couriers involved are assumed to either belong to a single carrier or provide crowdsourcing service to that carrier who centrally plans and manages all pickup and/or delivery services to its customers.

In spite of the different variants of urban transshipment problems, current transshipment literature considers a traditional hub-and-spoke setting in which incoming goods are unloaded at a transshipment facility (e.g., a sorting center, a distribution center, or an urban consolidation center), sorted, and transferred by an outbound service fleet to the remaining of the shipment. The hub-and-spoke paradigm is the state-of-the-practice particularly well suited for inter-regional shipping and non-expedited urban deliveries. On the other hand, little has been studied about new and innovative transshipment strategies particularly in the context of the emerging ODSD delivery service. The proposed P2PT strategies intend to fill that gap in the literature. Thus this paper adds significant value to the relevant literature.

Due to the page limitation, this abstract only presents the general definition and model structure of a P2PT problem. The following will be presented as part of the study: (1) a heuristic solution for the simplest form of P2PT, labeled as P2PT-1, which considers the scenario where there is only one single inter-zonal ODSD service request; (2) the model formulation and solution method are presented for a more general form of P2PT, labeled as P2PT-M, in which multiple inter-zonal ODSD service requests appear at the beginning of the daily operation; (3) a real-time P2PT scenario (labeled as P2PT-RT) in which the inter-
zonal ODSD requests appear in real time; and (4) the cost comparison of P2PT-M with direct shipping, the most responsive and naïve solution.

2. P2PT Problem Definition and Formulation

2.1 P2PT Network Setting

In P2PT, it is assumed that the entire study area can be divided into an (R+1) x (R+1) grid (see Figure 2), where R is the user-defined maximum number of relays allowed in a single transshipment from its origin to its final destination. An (R+1) x (R+1) grid guarantees that an inter-zonal package originated from any one zone in the grid can reach its final destination zone within R relays. In practice, R is likely no greater than 3. Each grid defines a service zone Z pre-assigned to a single courier whose daily operation does not go beyond this zone. The shape of a grid can be a rectangle or a square. In reality, the study area may be of an irregular shape. A smallest (R+1)x(R+1) grid that contains the entire study area is then used. Moreover, each service zone may not possess a nice rectangular or square shape in practice. In fact, as can be seen later in the model formulation the (R+1)x(R+1) grid is a weak and convenient assumption for model illustration purposes.

A neighboring zone of a given zone Z is defined as any zone that shares an arc or a vertex with zone Z. For example, in Figure 2, zone (1,1) has three neighboring zones, zones (1,2), (2,1) and (2,2). The corresponding couriers are called neighboring couriers. Note that in P2PT, because a single courier is pre-assigned to a single zone and a single vehicle, the terms courier, zone, and vehicle are interchangeable in the context of P2PT.

Two types of relay points are assumed. If the two neighboring zones share an arc, the relay point between these two neighboring zones is always at the midpoint of the shared arc, e.g., Point A in Figure 2 between zone (1,1) and zone (1,2). If the two neighboring zones share a vertex, the relay point of these two neighboring zones is at the shared vertex, e.g., Point B in Figure 2 between zone (1,1) and zone (2,2). It is further assumed that there is a secured drop box located at each relay point that both neighboring couriers have access to - one may think of mailbox as one kind of drop box. The drop box has sufficient capacity to hold packages temporarily during the daily operation.

Two types of demand are considered in P2PT: intra-zonal (local) demand and inter-zonal demand. Intra-zonal demand is assumed pre-determined prior to the start of the pre-assigned courier’s daily operation and handled solely by that courier in his/her daily routine. It is a classical vehicle routing problem (VRP). Inter-zonal demand corresponds to the ODSD service requests that P2PT is designed for. Each ODSD service request consists of a pair of pickup and delivery tasks, in which the package is picked up in an origin zone and delivered in a destination zone that is different from the origin zone. Thus, a P2PT strategy is an inter-zonal transshipment strategy with relays among a subset of couriers. It is worth noting that if the pickup and delivery locations of an ODSD service request are within the same service zone then the P2PT strategy reduces to a classical VRP.
Consider an example shown in Figure 2. Suppose there is an inter-zonal ODSD service request \((P,D)\), where the pickup point \(P\) is in zone \((1,1)\) and the delivery point \(D\) is in zone \((2,3)\). There are six possible transshipment strategies: 

\[\{(1,1), (1,2), (1,3), (2,3)\}, \{(1,1), (1,2), (2,3)\}, \{(1,1), (1,2), (2,2), (2,3)\}, \{(1,1), (2,2), (1,2), (2,3)\}, \{(1,1), (2,2), (2,3)\}, \{(1,1), (2,2), (1,2), (2,3)\}\] and 

\[\{(1,1), (2,1), (2,2), (2,3)\}\]. Each of these strategies involves different couriers and relay points. Such a partnership is temporary and breaks when the relays are completed. Thus, the objective of P2PT-1 is to find the transshipment (relay) strategy that incurs the minimum total network travel time of those couriers involved in the relays.

Additional conditions and assumptions that the proposed P2PT must satisfy are:

1. All service vehicles are homogeneous in size and type;
2. All couriers start their daily operation at a single depot (home base) \((O)\) at time zero;
3. The total work hour limit for each courier is 8 hours;
4. Inter-zonal demand must be served within the same day operation;
5. No idling is allowed at stops;
6. There is no extended waiting time on an arc or at customer stops;
   - No time window constraint is considered for any intra- and inter-zonal customer demand; all demand is served within the 8-hour work hour limit.

### 2.2 Two-level P2PT Model Structure

P2PT strategies consist of finding the zones (couriers) who could do the transshipment and intra-zonal routing within those zones involved in the transshipment. Thus we formulate the P2PT problem in a two-level model structure (Figure 3). The upper level is an inter-zonal transshipment path search that determines the relay strategy among the zones. Its outcome includes the zones and the relay points involved in the transshipment. They are the input to the lower level model, which is a P2PT-1 model that constructs the intra-zonal routing among the intra-zonal customers, inter-zonal ODSD customers (for origin and destination zones), and relay points for all those zones involved in the transshipment.

![Figure 3. Two-level P2PT model structure](image)

### 3. Solution Method and Results

An adaptive boundary relaxation (ABR) heuristic algorithm has been proposed and evaluated for its solution performance. It is shown to have comparable solution quality with the exact solution method, and huge computation time savings, requiring only 0.1% of computation time by the exact solution method.

The numerical results suggest that:

1) ATT of a single inter-zonal ODSD service request is highly volatile depending on the relative location of the service request to that of the intra-zonal customers;
ATT decreases as the demand for P2PT service (the economy of scale) increases, and P2PT strategies become much more cost efficient; and

3) as the economy of scale for ODSD service increases ATT tends to stabilize (converge) fairly quickly.

Our investigations have also found that the efficiency of P2PT, measured in terms of the additional travel time (ATT) incurred per inter-zonal ODSD request (hours/request), improves initially with the economy of scale, and then seems to flatten out in the static demand scenario or worsen in the real-time demand scenario as the number of inter-zonal.

The P2PT strategies have also been compared with the most responsive and naive delivery strategy of direct shipping. It is found that while P2PT tends to incur longer travel time than direct shipping, it provides the ODSD service without having to expand and maintain a larger fleet of vehicles (and thus drivers) or incurring extra labor cost. As the economy of scale of ODSD demand increases, those advantages of P2PT will only become more prominent. Because the underlying idea of P2PT is better utilizing vehicle capacity.

4. Future Work

There are many possible extensions of this research. One interesting extension is to incorporate simultaneous competition and collaboration among multiple couriers or even carriers. For example, at present, most carriers collect freight requests from shippers and then optimize the vehicle routing individually. However, carriers may benefit from collaboration by sharing resources (e.g., vehicle capacity) and labor cost more efficiently and effectively (Liu et al. 2010, Li 2013).

Another future research direction is use of electric vehicle (EV) in P2PT. EV has a range limit, which restricts EV's service range. Through P2PT, the service range of EV can be significantly extended even without the need for en-route battery recharging. In addition, time-dependent speed profile and time window constraints should be incorporated to reflect the true urban logistics operating environment.

5. Selected Reference


Balancing Availability and Profitability in E-Fulfillment with Revenue Management and Predictive Routing

Catherine Cleophas, RWTH Aachen University
Jan Fabian Ehmke, European University Viadrina
Charlotte Köhler, European University Viadrina
Magdalena A.K. Lang, RWTH Aachen University

The ongoing boom in e-commerce comes with increasing competition and logistic challenges, in particular for goods that require attended deliveries such as selling fresh groceries online. To enable attended deliveries, customers and e-grocers must agree on a delivery time window. To that end, after customers have assembled their virtual shopping baskets, e-grocers usually offer a selection of delivery time windows that customers can choose from.

On the one hand, customers’ expectations are high: they usually look for a tight delivery time window in the very near future, and they are not willing to pay significantly more for a better choice of delivery time windows. On the other hand, e-grocers’ profit margins are small, and the offered delivery time windows strongly affect the costs for order delivery. Hence, delivery time windows are challenging to plan and costly in execution. Optimizing solely for profitability, the provided selection of delivery time windows might be privileged to customers with high short-term revenues, ignoring important customer characteristics such as their social impact. We investigate how e-grocers can ensure the availability of delivery time windows not only for the most short-term profitable customers, but how they can balance the availability of delivery time windows for different customer segments while, at the same time, keeping costs of delivery under control.

To this end, we present a solution approach that focuses on profitability, but also ensures a sufficient availability of delivery time windows per delivery area and customer segment. We predict orders and their expected value, expected delivery cost, and customers’ expected social impact. We propose an iterative solution approach: First, predictive routing computes the allocable number of deliveries per time window and delivery area based on characteristics of forecasted order requests. Subsequently, a dynamic program determines the optimal set of delivery time windows to offer to each incoming order request. A final routing determines the minimum cost-routes of delivering the accepted orders. We evaluate the proposed approach in a virtual simulation laboratory modeling the e-fulfilment process. This simulation environment is empirically calibrated using data obtained from an industry partner.

Conceptual Framework

The process of fulfillment for e-commerce deliveries (e-fulfillment) spans three steps: (1) order capture and promise, (2) order sourcing, and (3) order delivery (Campbell and Savelsbergh 2005). Order capture aims to maximize the number or value of accepted orders. Given heterogeneous
demand as well as accurate customer segmentation and forecasts, planners can apply sophisticated methods from the domain of revenue management in this step. In this regard, Agatz et al. (2013) differentiate static versus dynamic allocation of delivery time windows as well as dynamic pricing of deliveries. Subsequently, order sourcing assembles the accepted orders, and the final step of order delivery aims to minimize the cost of delivering accepted orders given tight delivery time windows. Finding a cost-minimal route considering time-window constraints can be solved through the vehicle routing problem with time windows (VRPTW) as it is presented by Ehmke and Campbell (2014), for example.

Early contributions on e-fulfillment aim to minimize expected delivery costs while maximizing the number of accepted orders, but account neither for order value nor for the availability of delivery time windows. For example, Campbell and Savelsbergh (2006) use delivery fees to incentivize customers to choose specific delivery time windows. Alternatively, Agatz et al. (2011) control the availability of delivery time windows per zip code to minimize expected delivery costs. Ehmke and Campbell (2014) compare different order acceptance mechanisms to maximize the number of accepted orders while ensuring reliable deliveries considering time-dependent travel times.

Recent contributions mostly implement order capture by combining efficient routing with dynamic delivery fees. For example, Yang et al. (2014) and Yang and Strauss (2016) present an integrated model to compute optimal delivery fees. Following a similar idea, Yang et al. (2015) apply a differentiated time window pricing approach. However, as customers are sensitive to shipping charges, these can become a competitive disadvantage. In addition, given a limited span of acceptable delivery fees, the expected gains from optimizing this fee are relatively small. As one of the first to consider order value as a decisive factor of e-fulfillment, Cleophas and Ehmke (2014) propose to reserve delivery capacity for the most valuable expected orders.

However, none of these contributions consider customer loyalty and satisfaction when offering delivery time windows, although they are an important factor of retailers’ long-term success. When purely focusing on short-term profit, certain delivery time windows may never be offered in certain areas or to certain customer segments. Such a priori limiting of the delivery capacity for some delivery areas can systematically censor the demand forecast. When considering long-term objectives, systematically discriminating some customer segments may not be desirable, either. For example, some low-revenue customers may only try out the service and would, after positive experiences, increase their spending. Others may have a large social impact, spreading their – positive or negative – experience via social networks and thereby affecting potentially more valuable customer segments. In addition, when some areas are rarely catered to, potential marketing effects from the visibility of branded trucks are forfeit.

**Considering Profitability in Predictive Routing**

Based on a set of forecasted order requests, we introduce a predictive routing that computes the delivery capacity per time window and delivery area. In particular, we create routes that service the most valuable subset of forecasted requests for a given fleet of delivery vehicles. Given these objectives and constraints, we model the problem as Orienteering Problem with Time Windows.
The OPTW combines the well-known vehicle routing problem with time windows (VRPTW) with the Knapsack Problem to determine the most valuable set of order requests finding cost-minimal routes for a given fleet of vehicles while considering time window constraints (Vansteenweegen et al. 2011). Thus, the OPTW can be used to balance customer value and costs of delivery to minimize routing costs and maximize profits. Each customer resides in one delivery area. Therefore, we compute the expected travel time per resulting delivery from averages within and between areas. In addition, the expected travel time is adjusted to consider delivery area characteristics such as the expected density of customers or the density of the road network.

By extending the standard variant of the OPTW, we also consider minimum availability requirements in predictive routing. To that end, we prescribe a minimum number of deliveries to offer per time window and delivery area. To ensure that the predictive routing adheres to such minimum requirements, we complement the expected requests by an appropriate set of highly valuable pseudo-requests per area and time window. We will investigate several heuristics to solve the OPTW and compare the results of predictive routing with regard to availability of time windows, profitability of customers, and run time of solution algorithm.

From the determined set of expected orders and the resulting routes, we derive the number of deliveries per time window and delivery areas. These constitute the expected delivery capacity given a profitable routing for expected orders.

**E-Fulfillment Control**

E-grocers can deliberately control the offered delivery time windows to maximize the value of delivered orders. When maximizing short-term revenue, the e-grocer decides which delivery time windows to offer to a customer based on the shopping basket value. Customers may flexibly choose from a set of offered time windows per their preferences. We model time window choice through a multinomial logit model. To maximize revenue from orders, we present a dynamic program. For a given order request, this program determines the optimal set of time windows to offer, given the left-over delivery capacity as well as the value and time window choice of future expected requests.

To account for criteria beyond short-term revenue, we present a weighted indicator of order value. To exemplify this, we calculate the normalized revenue associated with the order’s shopping basket and combine this with a given normalized indicator of the associated customer segment’s expected social impact. Based on this weighted value, the dynamic program can optimize the time window offer set without neglecting the long-term value of customer segments.

**A Virtual Laboratory for E-Fulfillment Planning**

To evaluate the effect of balancing profitability and availability of time windows, we present a virtual laboratory for e-fulfillment planning. This laboratory forms the basis for a computational study evaluating the revenue and cost from accepted orders, given a set of customer segments and delivery areas. Our analysis particularly illustrates the resulting balance between providing a
minimum level of service across delivery areas, revenue earned, delivery cost, and the resulting availability experienced by the customers from different customer segments.

Figure 1: Virtual Laboratory

Figure 1 illustrates the composition of the virtual laboratory. To ensure an empirical foundation for experimental settings, the simulation’s parameters are calibrated using transactional data. The idea is to embed the e-fulfillment process in a framework of complex traffic and customer behavior. The e-fulfillment system is encapsulated and independent; thus, it could be extracted for a stand-alone implementation as part of an empirical case study.

To calibrate customer demand, we rely on transactional data provided by a major German e-grocer, DHL-owned Allyouneed Fresh (www.allyouneedfresh.de). Allyouneed Fresh offers fresh groceries online and delivers them in two-hour time windows in metropolitan areas. Figure 2 visualizes two insights of the available customer and order data. The left panel indicates the share of deliveries per time window, highlighting that the demand concentrates on a few time windows such as the morning and after work time windows. The right panel depicts the share of shopping basket values, highlighting that low-value baskets make up more than 50% of the orders.

Figure 2: Analysis of Order Data from an E-Grocer

From the empirical data, we derive a parametrization of customers’ preferences for select time windows. In addition, the indicated high variation in demand and basket values reflect the heterogeneous demand that is a prerequisite for successful revenue maximization during order capture.

Conclusion
To iteratively optimize the balance of delivery cost and order revenue for e-fulfilment, we propose to combine predictive routing and a dynamic program for time window offer set optimization. Thereby, we present an approach to maximize the value of acceptable deliveries, which also accounts for availability requirements and the expected long-term value of customer segments. We also present an empirically calibrated virtual laboratory to model e-fulfillment problems as well as results from a computational study set in this laboratory.

Future research is required to fully integrate routing and time window allocation. Such an integration would break up the iterative process of predictive routing and order acceptance to account for the opportunity cost of each newly accepted request given an adjustable delivery capacity. Finally, the research presented here could be transferred to further attended services such as medical care.

References


1 Introduction

With the rise of e-commerce, consumer preferences have been shifting significantly and accordingly the services offered to the market. Particularly, the last mile delivery services seem to act as key differentiators among e-commerce players. McKinsey recently published a report which states 25% of consumers are willing to pay some premium for a same-day delivery (see Joerss et al. (2016)). The rapid raise of various start-ups which offer different types of same-day delivery services is another indicator of the shift in the current last mile delivery environment. However, the design of a same-day delivery network brings new challenges such as high installation and operational cost.

Same-day delivery services that are the focus of this paper are characterized by two main properties. Firstly, all delivery requests arrive on within the same day of the service; i.e., a service provider has no a-priori information of any delivery task. Secondly, each request associates with a unique single product, which means that not any task can be substituted with another one. This turns the problem into a multi-commodity problem. In this particular problem setting, we assume that all delivery requests are satisfied from a single depot with a fleet of dedicated vehicles. The vehicles are loaded and dispatched from the depot. Each vehicle can carry multiple packages at the same time and can make multiple trips within the service period. The objective is to minimize the size of the fleet by delivering all arrival delivery requests. When there is a tie in the number of vehicles, the total travel time is the tie breaker. In the rest of the study, we refer to this problem as the same day delivery problem (SDDP).
Even though the same day delivery problem is inherently a dynamic problem, in this study we will focus on the static version of the problem. The aim of concentrating on the static version is two-fold. Firstly, subproblems of the same day delivery problem - the problem that corresponds to a single time epoch of the service period - is this static problem. Secondly, solving the same day delivery problem in hindsight provides a benchmark solution to the real life problem. The quality of the same-day delivery solution methods can be measured by this hindsight benchmark.

This research effort is to investigate the same day delivery setting described informally above. In particular, we formulate the static of the same day delivery problem (SSDP), i.e, all delivery requests are assumed to be known at the beginning of the day. Thus, the problem can be considered as a variant of the vehicle routing problem (VRP). Nevertheless, the problem differentiates itself from the canonical vehicle routing problem in several aspects. First of all, all delivery requests are not available for to be pickup at the depot at the beginning of the day. Each one can only be picked up after its own announcement (release) time. Secondly, vehicles can make multiple trips throughout the service period.

The complexity of the vehicle routing problems with release dates have been recently discussed in Archetti et al. (2015). Later on, Reyes et al. (2016) also introduce the the individual deadlines for each customer, and analyze the structural properties of the optimal schedules. Cattaruzza et al. (2016) are the first authors that combined the multi-trip vehicle routing problem with the routing problems with release dates. They introduce a problem which exists in the city logistics setting where two layers of delivery system has to be synchronized. Furthermore, Klapp et al. (2016) and Voccia et al. (2015) investigate the dynamic and the stochastic variant of the same day delivery problem. Regarding these studies, we can summarize the contribution of this paper are as follow.

1. We formulate the SSDP to capture the basic aspects of the batching and routing decisions for the same-day delivery. Furthermore, this formulation provides a theoretical benchmark for the dynamic and stochastic variant of the same day delivery problem.

2. We develop several heuristic approaches to determine multiple initial solutions and propose problem specific local operators for an adaptive large neighborhood search method.

3. We conduct extensive computational experiments to validate our proposed heuristics.
2 Problem Formulation

The same-day delivery problem (SDDP) can be defined on a complete undirected graph $G = (N, A)$, where $N$ is a set of nodes and $A$ is a set of arcs. Set $N = \{0, 1, \cdots, n\}$ consist of the destinations of all delivery tasks $P$, and the depot, denoted by node 0. $c_{ij}$ and $t_{ij}$ denote respectively the amount of distance and time if a vehicle traverses on arc $(i, j) \in A$. Each delivery task $i$ arrives and turns to a delivery order at its announcement time $a_i$ in the daily service period, $[0, T]$. Announcement time $a_i$ is the first time that the service provider gets informed about the product. After non-negative amount of preparation time $p$, task $i$ is ready to be picked up at its release time $r_i := a_i + p$. Each task $i$ has an individual delivery time window $[e_i, l_i]$, where $r_i \leq e_i < l_i$ (see Figure 1). $\bar{l}_i := l_i - t_{0i}$ denotes the latest time that the parcel has to leave the depot to be shipped on time.

In the SDDP, a solution composes of set of trips. A trip is a sequence of customer visits, which starts at the depot, and ends at the depot again. Each trip has to be assigned to one of the dedicated vehicles. Also, a vehicle can make multiple trips throughout the service period. A solution example of the SSDP can be found in Figure 2.

Figure 1: Timeline of a task

Figure 2: Multi-Trip Structure

An important aspect of this problem is the fact the each task has an individual timeline. This characteristic has an significant impact on the formation of trips and departure time of vehicles. A trip can only start after the maximum release time of all the tasks which are assigned to the trip. Since the selection of the task set determines the earliest departure of the trip, this time and the selection affects to route construction and consequently the overall solution quality. Furthermore, consecutive trips of the same vehicle have impact on each others. More specifically, a trip can only
start after the predecessor trip finishes, earliest and latest departure times of trips depends on each other.

The primary objective of the same day delivery problem is to minimize the number of vehicles that is required to serve all tasks, and the secondary objective is to minimize the total service duration. Furthermore, the following conditions should be satisfied in a feasible solution.

1. All tasks should be served within their time-windows.
2. Each task has to be assigned to exactly one trip.
3. Consecutive trips should start after the previous one finishes.
4. A trip does not start earlier than the maximum of the release times of the assigned tasks.
5. The total demand of the tasks in a trip does not exceed the vehicle capacity.

3 Solution Approach

As its core, the problem described as a same day delivery problem is a variant of vehicle routing problem with time windows. These problems are infamously known as hard problems, thus there are many heuristics have been proposed to solve them in reasonable time. In our solution approach, we develop an adaptive large neighbourhood heuristic (ALNS) with the problem specific removal and insertion operators other than the canonical ALNS operators, which is also a heuristic. Particularly, we introduce split-merge operators which take the multiple trips as an input and split them into smaller ones. Then, destroyed trips merge with others. The significant distinction of these operators is tackling with the multi-trip structure. Split and merge operators can be implemented on the trips that are in the same vehicle -intra-vehicle-, and that are assigned to the different vehicles inter-vehicle. Furthermore, we develop several constructive heuristics to create initial solutions. These heuristics can be briefly summarized as follow. The first heuristic that has been used is the basic greedy insertion. The others are combination of two basic operations: ranking and insertion. For ranking phase, we used two different approaches. First one is temporal ranking, which means the delivery tasks are ordered according to their release times. The second ranking comes from solving a variant of the traveling salesman problem. For the insertion phase, we use sequential insertion which means that a vehicle is loaded with trips until no more load is possible.
To evaluate the performance of our approach we use a set of instances. Rather than adapting the well known instances, we create an 15 km square plane in which the depot is located at the middle. The location of the delivery tasks are distributed uniformly and randomly over the plane. The announcement times of delivery tasks are randomly drawn from 9 am to 7 am. Some preliminary results that are obtained from the several constructive heuristics be seen in Table 1. This table reveals the number of necessary vehicles, and trips to satisfy all task deliveries. The last column gives the average number tasks per trip.

<table>
<thead>
<tr>
<th># nodes</th>
<th>Vehicles</th>
<th>Trips</th>
<th>Trips per vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chrono</td>
<td>TSP</td>
<td>Greedy</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>150</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>150</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>200</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>200</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 1: Vehicle and Trip Numbers

References


Supply Chain Logistics & Methods
TD4: Service Network Design
Thursday 4:30 – 6:00 PM
Session Chair: Mark Hewitt

4:30 A Benders’ Decomposition Approach for Airline Timetable Development and Fleet Assignment
Keji Wei*, Vikrant Vaze
Thayer School of Engineering-Dartmouth College

5:00 Service Network Design of Bike Sharing Systems: Formulation and Solution Method
1Bruno Albert Neumann Saavedra*, 2Dirk Mattfeld, 2Teodor Gabriel Crainic, 2Bernard Gendron, 4Michael Römer
1Technische Universität Braunschweig, 2ESG, UQAM & CIRRELT, 3Université de Montréal, 4Martin Luther University Halle-Wittenberg

5:30 Enhanced Dynamic Discretization Discovery Algorithms for Service Network Design Problems
Mike Hewitt*
Loyola University Chicago
A Benders’ Decomposition Approach for Airline Timetable Development and Fleet Assignment

Keji Wei, Vikrant Vaze

Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755
{keji.wei.th@dartmouth.edu, vikrant.s.vaze@dartmouth.edu}

Keywords: schedule design; fleet assignment; valid inequalities; Benders’ Decompositions

Abstract:
In this paper, we describe a model that integrates timetable development and fleet assignment steps from the airline planning process to maximize airline profits. We attempt to select the optimal set of flight departure times and the optimal fleet type for every flight leg. The optimal solution captures a large market share by explicitly incorporating passengers’ itinerary choice, while ensuring that the operating cost stays low. We design a novel way to apply Benders’ decomposition to solve this challenging mixed-integer optimization problem to near-optimality. Computational results using data from two real-world airline networks demonstrate the efficacy of the proposed modeling and solution techniques.

Introduction:
Typical airline planning process consists of schedule design (also known as timetable development), fleet assignment, aircraft routing and crew scheduling. Across these four steps, airline timetable development and fleet assignment steps have the largest effect on the airline operating costs as well as on the airline’s revenue. There is an expanding literature seeking to improve airline schedules by integrating some stages in the airline planning process. In particular, Sherali et al. (2013) integrate fleet assignment model with some incremental schedule design steps by selecting optional legs and assigning aircraft types to each legs.

Compared to Sherali et al. (2013), our paper has the following contribution. First, our model is the first comprehensive timetable development effort in the sense that flight departure times are fully flexible. Second, the market shares for the host airline’s itineraries are adjusted automatically according to changes in each flight’s timetable.

Model Formulation:
In this section, we first present the mathematical notation used in our work.

Sets
OA: Set of other airlines (and with the no-fly alternative included).
A: Set of airports indexed by $a$.
$M$: Set of all markets indexed by $m$.
$I_m$: Set of all itineraries in market $m$ indexed by $i$.
$L$: Set of all segments indexed by $l$. 
\( T \): Set of all time slots in a day, indexed by \( t \).
\( F \): Set of all fleet types indexed by \( f \).
\( N \): Set of nodes in the timeline network, indexed by \((f, a, t)\), for fleet type \( f \), airport \( a \), and time \( t \).
\( N' \): Subset of nodes in the timeline network excluding the last node for every airport-fleet type combination.
\( \text{In}(f, a, t) \): Set of inbound flight legs into node \((f, a, t)\).
\( \text{Out}(f, a, t) \): Set of outbound flight legs from node \((f, a, t)\).
\( I / I_{NS} / I_{OS} \): Set of all itineraries/nonstop itineraries/one-stop itineraries that could be potentially offered by the airline.
\( I = I_{NS} \cup I_{OS} \).
\( I_{NS} = \{l, t | l \in L, t \in T\} \): Set of nonstop itineraries (same as flight legs) indexed by segment \( l \in L \) and departure time \( t \in T \).
\( I_{OS} = \{l, k, t, v | l \in L, f \in F, t \in T\} \): Set of one-stop itineraries indexed by the segment \( l \in L \) and the departure time \( t \in T \) of the first flight and the segment \( k \in L \) and the departure time \( v \in T \) of the second flight.
\( I_{m}(l, t) \): Set of all (nonstop and one-stop) itineraries serving market \( m \) which use a flight in segment \( l \) with departure time \( t \).

**Parameters**
\( \text{Dem}_m \): Total demand in market \( m \).
\( \text{frequency}_l \): Number of flights per day on segment \( l \).
\( p_i \): Ticket price of itinerary \( i \).
\( a_i \): \( e^{Ui} \), Attractiveness of itinerary \( i \), (see Coldren and Koppelman 2005).
\( a_0 \): Sum of attractiveness of all itineraries of other airlines, and of the no-fly alternative, in market \( m \).
\( \text{ope}_{l,t,f} \): Operating cost of a flight in segment \( l \) and departure time \( t \), operated by fleet type \( f \).
\( \text{Avaf} \): Available number of aircraft of fleet type \( f \).
\( \text{minE}_{-a} \): Time just before the first event at airport \( a \).
\( \text{Cap}_f \): Seating capacity of an aircraft of fleet type \( f \).

**Decision Variables**
\( s^m_0 \): Total market share of all itineraries of the other airlines, and of the no-fly alternative, in market \( m \).
\( s_i \): Market share of itinerary \( i \) in its corresponding market.
\( x_{l,f,t} = 1 \) if fleet type \( f \) is assigned to a flight in segment \( l \) with departure time \( t \), 0 otherwise.
\( y_{f,a,t} / y_{f,a,t}^- \): Number of aircraft of fleet type \( f \) on ground at airport \( a \) just before/after time \( t \).

Accordingly, we formulate integrated schedule design and fleet assignment model (SF) as follows.

Maximize:
\[
\sum_{m \in M} \left( \text{Dem}_m \sum_{i \in I_m} p_i s_i \right) - \sum_{l \in L} \sum_{f \in F} \sum_{t \in T} \text{ope}_{l,t,f} x_{l,f,t}
\]

Subject to:
\[
\sum_{m \in M} \left( \text{Dem}_m \sum_{i \in I_m} (l,t) s_i \right) \leq \sum_{f \in F} (\text{Cap}_f \times x_{l,f,t}) \quad \forall l \in L, t \in T
\]
\[
s^m_0 + \sum_{i \in I_m} s_i = 1 \quad \forall m \in M
\]
\[
a_0^m s_i \leq a_i s_0^m \quad \forall i \in I_m, m \in M
\]
\[
\sum_{t \in T} \sum_{f \in F} x_{l,f,t} = \text{frequency}_l \quad \forall l \in L
\]
\[
\sum_{f \in F} x_{l,f,t} \leq 1 \quad \forall l \in L, t \in T
\]
\[
\begin{align*}
\Sigma_{f \in F} x_{l,f,t} & \geq s_i \quad \forall \; i \in I_m (l,t), m \in M, l \in L, t \in T (7) \\
\Sigma_{a \in A} y_{f,a} \cdot min_{E_a} & \leq \text{Av}_{l} \quad \forall \; f \in F (8) \\
y_{f,a,t} + \Sigma_{(l,f,t) \in I_m (f,a,t)} x_{l,f,t} & = y_{f,a,t+} + \Sigma_{(l,f,t) \in \text{out}(f,a,t)} x_{l,f,t} \quad \forall \; (f,a,t) \in N (9) \\
y_{f,a,t+} = y_{f,a,t} & \quad \forall \; (f,a,t) \in N' (10)
\end{align*}
\]
\[
\begin{align*}
x_{l,f,t} & \in \{0,1\} \quad \forall \; l \in L, f \in F, t \in T (11) \\
y_{f,a,t}, y_{f,a,t+} & \in \mathbb{Z}^+ \forall \; (f,a,t) \in N (12) \\
s_i & \geq 0 \quad \forall \; i \in I (13)
\end{align*}
\]

The objective function (1) is to maximize the profit as given by revenue minus the cost. Constraint (2) ensures that the number of passengers on each flight leg does not exceed the capacity of the aircraft type assigned to that flight leg and is zero if no aircraft type is assigned to that flight leg. Market demand constraint (3) guarantees that the total accepted demand of all itineraries (including the itineraries of host airline, itineraries of the other airlines, and the no-fly alternative) in a specific market must be equal to the total demand of that market. Demand-splitting constraint (4) ensures that the spilled demand is recaptured in such a way that the market share of the itinerary is at most proportional to the attractiveness of the itinerary. Constraint (5) ensures that the total number of flight legs operated on each segment is equal to the segment’s flight frequency as determined during the frequency planning stage. Constraint (6) stipulates that not more than one aircraft type can be used to operate a flight on a specific segment at a particular time. Constraint (7) ensures that an itinerary is not feasible if any leg of the itinerary is not operated. Constraints (8), (9) and (10) represent flow balance and aircraft availability restrictions (see Sherali et al. (2013), for example). Finally, constraints (11), (12), (13) are variable value constraints on the decision variables.

Solution Approach

Although the \((y_{f,a,t}, y_{f,a,t+})\) variables, representing the number of aircraft on ground arcs, are required to be integer-valued, we can declare them as continuous and \((y_{f,a,t}, y_{f,a,t+})\) are automatically integer-valued for any given values of binary variables \(x_{l,f,t}\). So, we can relax the integrality restriction on the \((y_{f,a,t}, y_{f,a,t+})\) variables in our model and denote resulting model as SF+.

We next apply Benders’ decomposition to model SF+ by formulating a master program that includes only the binary variables \((x_{l,f,t})\) and by letting other variables \((y_{f,a,t}, y_{f,a,t+}, s_i)\) define the sub-problem (Sherali et al., 2013). That yields the following decomposition:

Master problem

Maximize: \(\hat{Z} = \Sigma_{l \in I} \Sigma_{f \in F} \Sigma_{t \in T} \text{opt}_{l,f,t} x_{l,f,t}\) \quad (14)
\[
\begin{align*}
\Sigma_{t \in T} \Sigma_{f \in F} x_{l,f,t} & = \text{frequency}_l \quad \forall \; l \in L (15) \\
\Sigma_{f \in F} x_{l,f,t} & \leq 1 \quad \forall \; l \in L, t \in T (16) \\
x_{l,f,t} & \in \{0,1\} (17) \\
a^{(k)} f - \hat{Z} & \leq b^{(k)} \quad \forall k \in 0 (18) \\
a^{(k)} f y & \leq b^{(k)} \quad \forall k \in F (19) \\
\hat{Z} & \geq 0 (20)
\end{align*}
\]
where \( \hat{Z} \) is a surrogate for the component \((z)\) of the subproblem’s objective function value. Constraints (17) and (18) are Benders’ cuts that are derived from the dual of the subproblem. \( a^{(k)} \) and \( b^{(k)} \) respectively are the coefficient vector and constant term for a cut generated during the dual solution of the subproblem, \( O \) is the index set of “optimality” cuts and \( F \) is the index set of “feasibility” cuts.

**Subproblem**

Maximize: \[ z = \sum_{m \in M} (Dem_m \sum_{i \in \mathcal{I}_m} p_i s_i) \]  

\[
\sum_{m \in M} \left( Dem_m \sum_{i \in \mathcal{I}_m (l,t)} s_i \right) \leq \sum_{f \in F} \left( Cap_f \cdot x_{l,f,t} \right) \quad \forall l \in L, t \in T \quad (\eta_{lt})
\]

\[ s_o^m + \sum_{i \in \mathcal{I}_m} s_i = 1 \quad \forall m \in M \quad (\lambda_m) \]

\[ A^m s_i \leq A_is_o^m \quad \forall i \in \mathcal{I}_m, m \in M \quad (24) \]

\[
\sum_{f \in F} x_{l,f,t} \geq s_i \quad \forall i \in \mathcal{I}_m (l,t), m \in M, l \in L, t \in T
\]

\[
\sum_{a \in A} y_{f,a} \cdot \min_{E_a} \leq A\text{val}_f \quad \forall f \in F \quad (\beta_f)
\]

\[
y_{f,a,t}^- + \sum_{(l,f,t) \in \mathcal{I}_m (f,a,t)} x_{l,f,t} = y_{f,a,t}^+ + \sum_{(l,f,t) \in \text{out}(f,a,t)} x_{l,f,t} \quad \forall (f,a,t) \in N
\]

\[
y_{f,a,t^-} = y_{f,a,t}^+ \quad \forall (f,a,t) \in N
\]

\[
s \geq 0
\]

(29)

Note that the variables \( x_{l,f,t} \) are treated as constants in the subproblem. The optimal dual variable values \((\eta_{lt}, \lambda_m, \beta_f)\) obtained by solving the subproblem are used to generate the Benders’ cut (31-32) to add to Master problem as constraint (18-19) as follows.

\[
\hat{Z} \leq \sum_{m \in M} \lambda_m + \sum_{f \in F} \beta_f \text{Aval}_f + \sum_{l \in L} \sum_{t \in T} \left( \sum_{f \in F} \left( Cap_f \cdot x_{l,f,t} \right) \right) \eta_{lt}
\]

\[
0 \leq \sum_{m \in M} \lambda_m + \sum_{f \in F} \beta_f \text{Aval}_f + \sum_{l \in L} \sum_{t \in T} \left( \sum_{f \in F} \left( Cap_f \cdot x_{l,f,t} \right) \right) \eta_{lt}
\]

**Computational Experiments**

CPLEX 12.7 solver with its default settings is used to solve all formulations above. An 8-thread / 4-core Intel® i7-XS6000 CPU with 8GB RAM and Windows 7 Professional as the operating system was used for all computational experiments. We performed our preliminary experiments using a small dataset based on a mid-size US-based airline. We assumed a time discretization of 15 minutes. Table 1 presents the results obtained using a 15-hour computational time limit for solving each problem instance. The gap is the difference between the current best integer objective function and the objective function of the best remaining node, taken as a percentage of the objective function of the best remaining node. As evident from Table 1, our model can solve instances involving less than 200 flight networks in a limited time and optimality gap is satisfactory, when compared with past studies that conduct incremental schedule design.

These promising early results have motivated us to pursue the Benders’ decomposition approach further in an attempt to solve for larger networks. As our next steps, we will now use this approach,
combined with other computational tricks such as valid inequalities, lifting, variable fixing, greedy search of branch-and-bound tree, etc. to attempt to solve larger scale problems involving major US-based airlines.

Table 1 Computational Results with Different Number of Airports

<table>
<thead>
<tr>
<th>No. of Airports</th>
<th>Best node</th>
<th>IP Solution</th>
<th>Computational time (hours)</th>
<th>Time slots for airport departure</th>
<th>Gap</th>
<th>No. of Flights</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>731142.2158</td>
<td>658359.9</td>
<td>15</td>
<td>15 minutes</td>
<td>11%</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>1019588.455</td>
<td>899522.5</td>
<td>15</td>
<td>15 minutes</td>
<td>13.35%</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>1673829.866</td>
<td>1540266.3</td>
<td>15</td>
<td>15 minutes</td>
<td>8.67%</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>1869874.461</td>
<td>1707531.3</td>
<td>15</td>
<td>15 minutes</td>
<td>9.51%</td>
<td>96</td>
</tr>
<tr>
<td>7</td>
<td>2107817</td>
<td>1896185.2</td>
<td>15</td>
<td>15 minutes</td>
<td>11.16%</td>
<td>106</td>
</tr>
<tr>
<td>8</td>
<td>2354654.548</td>
<td>2172203.2</td>
<td>15</td>
<td>15 minutes</td>
<td>8.40%</td>
<td>115</td>
</tr>
<tr>
<td>20</td>
<td>3733555.592</td>
<td>3562901.0</td>
<td>15</td>
<td>15 minutes</td>
<td>4.79%</td>
<td>171</td>
</tr>
</tbody>
</table>

Conclusions

In this paper, we optimize departure time and fleet assignment of each flight where the passenger demand for each itinerary is dependent on the itinerary attributes and attractiveness. A Benders’ decomposition based approach is developed to expedite the solution process.

As next steps, it would be interest to consider incorporating other variables, such as whether or not a flight leg is operated at a certain time of day, irrespective of the assigned fleet type, in our model to enhance the solution speed of the model by tightening its linear programming relaxation. In addition, further research is necessary for developing valid inequalities that are expected to enhance the solution speed as well.

References

1 Introduction

Station-based bike sharing systems (BSS) have become a viable, inexpensive, and sustainable complement to public transportation systems (Ricci, 2015). In particular, BSS offer one-way-capable bike trips, where each trip is usually free-of-charge within a time span of half-hour (Martinez et al., 2012). One-way bike trips are a suitable solution for mid-distance trips and the “last mile” between metro/bus stops and final destinations (DeMaio, 2009). BSS providers strive for establishing a high service level, that is, the percentage of users which successfully rent and return a bike in the desired stations when requested. Unfortunately, user trips lead to bike imbalances across the stations negatively affecting the level of service. To be more concrete, due to spatial and temporal characteristics of such bike trips and limited number of bike racks at stations, they regularly run full or empty within the day (OBrien et al., 2014). As consequence, users cannot find a bike to perform a desired trip or are unable to return their bikes afterwards. An important instrument for increasing the service level is to redistribute bikes among the stations during the day in such a way that users can perform their desired trips. However, since incomings generated by BSS are scarce due to the pricing model, resources for redistribution are limited to a certain budget.

The planning of bike redistribution involves decisions about inventory, transportation, and schedule. Inventory decisions determine the fill level at stations, that is, the ratio between the number of bikes currently located at the station and its capacity. Inventory decisions are made in a closed system, where bikes can be moved several times by vehicles and users. Transportation decisions define where and how
many bikes are picked-up or delivered at stations. Schedule decisions determine vehicle tours, where each vehicle tour is defined by a time-ordered sequence of stations visits, starting and ending such a tour at the depot. When deciding about the schedule, we need to note that drivers spend a significant amount of time on handling bikes at stations (de Chardon et al., 2016). Therefore, it is crucial to consider both vehicle travel times and handling times for loading and unloading bikes in order to obtain implementable redistribution plans. For real-world BSS, producing a suitable redistribution plan becomes a challenging task due to the large number of stations, complex usage patterns, and a limited resources to operate vehicles (Buttnér and Petersen, 2011). In addition, inventory, transportation, and schedule decisions are strongly interwoven parts of the redistribution problem. For instance, even a slight modification of one vehicle tour has an impact on the transportation decisions, which in turn affects the fill levels at stations.

In this work, we address the BBS redistribution problem at the tactical planning level. In particular, tactical planning level defines policies to operate vehicles efficiently while maximizing the service level to the extent possible by the limited resources. At the core of the tactical planning, the service network design of bike sharing systems (SNDBSS) selects regular redistribution services to satisfy mobility demand of users. These services take the form of “master tours” regularly operated each day over a certain season.

Our main contribution is to introduce a deterministic modeling framework for the SNDBSS adequately addressing the user demand, the handling of unmet demand, synchronization of redistribution and vehicle movements, as well as the representation of bike loading and unloading times and redistribution costs subject to a certain budget. In addition, we propose a solution framework capable of producing good-quality solutions in an acceptable time.

2 Problem Description

The underlying static network consists of two types of physical nodes: stations and the vehicle depot. Bike lanes and vehicle paths connect physical nodes accordingly. Each station has a rack capacity, whereas no bikes can be parked at the depot. The number of bikes in the BSS remain constant within the day, while each bike can be moved several times by both users and vehicles.

User demand estimation takes the form of user flows. Each user flow is characterized by an origin-destination station and timestamp pair. A user flow is successfully performed when at least one bike and free rack need to be ensured at the origin and destination station, respectively, at the requested times.

Vehicles are defined in terms of speed, load capacity, and handling capacity, defined as the time required for a driver to pick up or delivery a bike. Vehicles visit stations to realize handling operations according to their shortage of bikes or empty racks. That is, bikes are either picked-up at stations with shortage of racks or delivered at stations with shortage of bikes.

The objective is to minimize the total cost of using vehicles, operating vehicles, and handling bikes, together with the total cost of lost demand.
Figure 1: Time-expanded network.

Figure 1 illustrates an exemplary time-expanded network to represent time-dependent characteristics of the BSS redistribution problem. The y-axis represents physical nodes, whereas the x-axis represents points in time. Each circle illustrates a physical node replication at a point in time, where the black fill inside each circle represents the inventory at the corresponding node. The more black fill, the higher fill level at the node. Nodes are connected by different type of links. A user link permits user flows though it. User flows are performed if the fill levels at stations are adequate. Vehicle links allow the possibility of designing a holding, movement, or handling decisions, accordingly. By way of example, a master tour is depicted, following the node sequence $n_0 \rightarrow n_1 \rightarrow \ldots \rightarrow n_6$. The vehicle stays at the depot the first time period, represented by the holding link $n_0 \rightarrow n_1$. Links $n_1 \rightarrow n_2$, $n_3 \rightarrow n_4$, and $n_5 \rightarrow n_6$ allow vehicle movements among physical nodes and the transport of bikes between stations. A handling operation occurs in link $n_2 \rightarrow n_3$ representing that the vehicle stays at station 2 during the third time period. Since bikes were picked up at station 2, the fill level decreases. In the same way, $n_4 \rightarrow n_5$ represents delivering operations at station 3 for one time period, increasing thus the fill level at this station.

According to the time-expanded network previously defined, we embed the SNDBSS into a mixed-integer linear programming (MILP) formulation. Outputs are the fill levels at nodes (inventory), the bike flows performed by the master tours (trans-}

3 Matheuristic Solution Method

Our solution method combines exact and heuristic search techniques to construct a solution, making decisions hierarchically. The core of our hierarchical solution method are two optimization models. The first one is a dynamic transportation problem (DTP) which yields target fill levels and transportation services (Vogel et al., 2014). Transportation services are described by an origin-destination node
pair, and number of transported bikes. The DPT does not produce a feasible vehicle tour since it neglects the design of schedules and empty movements for the vehicle fleet. To produce feasible vehicle tours, we split each transportation service into a pick-up request and a delivery request. A delivery request can be satisfied by bikes provided by any pick-up request. Defining a time interval where such requests can be satisfied, the second problem to solve is a pick-up and delivery with time windows (PDPTW). PDPTW outputs are the design of vehicle tours, together with the (un-)satisfied requests.

Note, that both DTP and PDPTW are still MILP problems with a large number of integer variables. In the case of the DTP, we use heuristic techniques to limit the MILP exploration to a suitable number of integer variables, fixing to zero all integer variables which are not considered in the exploration. Thus, a MILP solver is capable of obtaining feasible solutions in an acceptable time. Our heuristic techniques are based on how bike redistribution is performed in practice, that is, transporting bikes from stations with high fill level to stations with low fill level. In the case of PDPTW, we adapt and use the efficient neighborhood-based metaheuristics extracted presented by Pisinger and Ropke (2010).

Before conducting a new iteration of our hierarchical approach, we feed the DTP with constraints derived from the PDPTW solution. We add constraints to the DPT imposing that transportation services depart from the area of the search space where a pick-up request were satisfied, arrive to the area of the search space where a delivery request were satisfied. In addition, we prohibit transportation services in the areas of the search space where requests were not satisfied. After adding such constraints, we conclude the iteration of our hierarchical approach by solving the DPT, obtaining new inventories.

The hierarchical approach is repeated until 1) vehicle tours do not significantly change, or 2) the service level cannot be improved within iterations. Vehicle tours obtained at each iteration are stored in a pool of promising vehicle tour. This pool is used as a warm start for CPLEX solver when the whole SNDBSS MILP problem is taken into account.

4 Computational results

In order to evaluate the performance of the proposed solution method, we generate small- and mid-sized instances up to 400 stations. The generation of instances is based on the online available data provided by four BSS programs located in the United States: Bay Area (San Francisco), Nice Ride (Minneapolis), Capital Bikeshare (Washington), and Citibike (New York City). We split the time horizon of one day into 5-minute intervals.

Experiments reveal that our matheuristic solution method can fulfill the expected demand for the small instances. In all conducted experiments our matheuristic outperforms the standard CPLEX MILP algorithm, which is not able to produce feasible solutions after 10 hours runtime.

Regarding the mid-sized instances, the matheuristic produces solutions with a level of service around 98%. Interesting, the consideration of additional redistribu-
tion vehicles does not lead to significant service level increments. Solutions reveal the existence of high-demanded stations where requests cannot be fulfilled even though several redistribution efforts. This is because these BSS programs offer valet service, that is, staff at such critical stations according to time-of-day in order to accept more bikes than the typically allowed. Incorporating valet service increases yield a level of service very close to 100%.

References


Enhanced Dynamic Discretization Discovery
Algorithms for Service Network Design Problems

Mike Hewitt, Loyola University Chicago

Consolidation carriers transport shipments that are small relative to trailer capacity. They participate in (1) the less-than-truckload (LTL) freight transport sector, a sector with annual revenues of about $30 billion, and (2) the small package/parcel transport sector, a sector with much larger annual revenues, with one player alone (UPS) reporting $54 billion in revenue in 2012. That said, both sectors are important, as both LTL and small package carriers play a prominent role in the fulfillment of orders placed online (as well as other channels). Fast shipping times (and low cost) are critical to the success of the online sales channel, with e-tailers such as Amazon.com continuously pushing the boundary, aiming for next-day and even same-day delivery. These trends result in increased pressure on consolidation carriers to deliver in less time (without increasing their cost). For example, for a large LTL carrier with whom we have collaborated, over 80% of their shipments need to be delivered within two days.

To deliver goods in a cost-effective manner, these carriers must consolidate shipments, which requires coordinating the paths for different shipments in both space and time. Short delivery times reduces the margin for error in this coordination, which necessitates planning processes that accurately time dispatches. These planning processes have long been supported by solving a variant of the Service Network Design (SND) problem (Crainic 2000, Wieberneit 2008), which decides the paths for the shipments and the services (or resources) necessary to execute them. Fundamentally, the design decisions for a consolidation carrier have both a geographic and temporal component, e.g., “dispatch a truck from Chicago, IL to Atlanta, GA at 9.05 pm.” The most widely-used technique for modeling the temporal component of these decisions is discretization; instead of deciding the exact time at which a dispatch should occur (e.g., 7.38 pm), the model decides a time interval during which the dispatch should occur (e.g., between 6pm and 8 pm).
With a discretization of time, service network design problems can be formulated on a time-expanded network (Ford and Fulkerson 1958, 1962), in which a node encodes both a location and a time interval, and solutions prescribe dispatch time intervals for resources (trucks, drivers, etc.) and shipments. That said, how a discretization is done impacts both the accuracy of the resulting SND model and its solvability. On the one hand, the more granular the discretization, the more accurate the estimation of consolidation opportunities, albeit at the expense of a larger network and harder optimization model to solve. On the other hand, a coarser discretization will yield a smaller network and easier optimization model, albeit one that is less accurate.

However, this trade-off between accuracy and model size is primarily a function of determining a discretization of time in a static, a priori manner, wherein one first determines a discretization of time, and then solves the resulting SND. Examples of such approaches can be found in Jarrah et al. (2009), Andersen et al. (2011), Erera et al. (2013a), Crainic et al. (2014). Boland et al. (2015) have instead presented a Dynamic Discretization Discovery (DDD) algorithm, which instead determines a discretization of time in a dynamic and iterative fashion, wherein at each iteration a SND is solved on a carefully constructed time-expanded network based upon the current discretization of time. To be precise, at an iteration and a given discretization of time, a partially time-expanded network is constructed in such a way as to ensure that the resulting SND yields a lower bound on the costs achieved if instead a SND was solved over a “full” discretization of time, wherein each location is represented at each point in time. We illustrate a flow chart of this type of approach in Figure 1.

To benchmark the performance of this algorithm for a desired discretization of time, Boland et al. (2015) compare it to solving the resulting SND on a time-expanded network based upon a full discretization of time. They perform this comparison for different discretizations of time, ranging from 60 minutes to just 1 minute. Relative to such an approach, the DDD algorithm is very effective; for

![Flow chart of a dynamic discretization discovery algorithm.](image-url)
each desired discretization of time, the DDD algorithm solves more instances in less time.

In this talk, we extend the research presented in (Boland et al. 2015) in both algorithmic and problem setting directions. Regarding the algorithmic development, DDD iteratively examines and then refines a partially time-expanded network until it can conclude that it has found one that will yield an optimal solution one could obtain if in fact a full enumeration was performed. The first enhancement we propose is a new technique for refining such a network (Step 4 in 1). However, we also show that the algorithm can be executed in a two-phased approach, wherein first it is applied to the linear programming relaxation of the SND, with the resulting time-expanded network used to seed the solution process for the integer version.

We report the results of employing these enhancements, for a desired discretization of time of five minutes, in Table 2. There, we report the average time to terminate (Time), the average optimality gap reported at termination (Opt. Gap), and the percentage of instances solved (Solved). We report results for the algorithm proposed in (Boland et al. 2015) (DDD), that algorithm when executed in a two-phased manner (DDD-2-phased), and that algorithm when executed in a two-phased manner and an enhanced method for performing step 4 (DDD-Enhanced). We also report these reports for the benchmark of solving an SND formulated on a time-expanded network based upon a full discretization of time (Full). We see that these enhancements improve the performance of the algorithm on all three performance measures.

 Regarding problem setting (Boland et al. 2015) propose their DDD algorithm, and prove its correctness, in the context of a classical SND model, one that models few of the operational realities carriers face. In this talk, we will also show how DDD can be adapted to many variants of the SND that are seen in practice without sacrificing computational performance. For example, in many operational settings, shipment paths must form a directed in-tree (Erera et al. 2013a, Jarrah et al. 2009, Powell 1986); namely, all shipments at a given terminal that have the same destination must travel to the same next terminal. We show how DDD can be adapted to this operational reality. Similarly, the SND considered in (Boland et al. 2015) required that

<table>
<thead>
<tr>
<th>Method</th>
<th>Time (sec.)</th>
<th>Opt. Gap</th>
<th>Solved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>2,573.24</td>
<td>5.62%</td>
<td>69.91%</td>
</tr>
<tr>
<td>DDD (Boland et al. 2015)</td>
<td>1,237.91</td>
<td>.92%</td>
<td>88.43%</td>
</tr>
<tr>
<td>DDD-2-phased</td>
<td>990.44</td>
<td>.77%</td>
<td>90.23%</td>
</tr>
<tr>
<td>DDD-Enhanced</td>
<td>847.40</td>
<td>.71%</td>
<td>91.90%</td>
</tr>
</tbody>
</table>

Figure 2  Impact of algorithmic enhancements for desired discretization of time of 5 minutes.
a shipment follow a single path from its origin terminal to its destination terminal. In this paper we show how DDD can applied to situations wherein a shipment can be split, and flow along multiple paths. Similarly, (Boland et al. 2015) focus on a SND that decides which arcs a shipment should flow on, and then, in turn, ensures that those arcs form paths. However, in some operational settings, potentially due to special consideration that must be paid to some shipments, the route of a shipment can be constrained to one of of a pre-defined set of paths (Erera et al. 2013b). We show how DDD can be adapted to this setting as well.

Thus, fundamentally, we believe the contribution of this talk will be that it will make DDD more relevant to practice in multiple ways. First, we will show how to improve its performance, so that it can solve more instances of the SND for a fine discretization of time, and in less time. Second, we will show how DDD can be applied to variants of the SND that represent a wide array of operational realities.

References

Andersen, Jardar, Marielle Christiansen, Teodor Gabriel Crainic, Roar Grønhaug. 2011. Branch and price for service network design with asset management constraints. Transportation Science 45(1) 33–49.


Erera, Alan, Michael Hewitt, Martin Savelsbergh, Yang Zhang. 2013b. Improved
load plan design through integer programming based local search. *Transportation Science* To appear.


Supplier Chain Logistics & Methods
FA4: Location Routing
Friday 8:30 – 10:30 AM
Session Chair: Francesco Viti

8:30  Finding Optimal Park-and-Ride Facility Locations in an Urban Network
     Pramesh Kumar*, Alireza Khani
     University of Minnesota

9:00  Reliable Facility Location Design with Imperfect Information: Continuum and Discrete Models
     Lifen Yun, Hongqiang Fan, Xiaopeng Li*
     University of South Florida

9:30  Location-Routing Problems with Economies of Scale
     James Bookbinder*, Xiaoyang Pi
     University of Waterloo

10:00 Exact and Approximate Optimal Route Set Generation in Sensor Location Problems
     Marco Rinaldi*, Francesco Viti
     University of Luxembourg
Finding optimal park-and-ride facility locations in an urban network

Pramesh Kumar  
Department of Civil, Environmental, and Geo-Engineering  
University of Minnesota Twin Cities  
500 Pillsbury Drive S.E., Minneapolis MN 55418  
Email: kumar372@umn.edu

Alireza Khani (Corresponding Author)  
Department of Civil, Environmental, and Geo-Engineering  
University of Minnesota Twin Cities  
500 Pillsbury Drive S.E., Minneapolis MN 55418  
Email: akhani@umn.edu  
Phone: +1 (612) 624-4411

ABSTRACT  
Park-and-ride facilities are becoming more popular nowadays as these facilities encourage people to switch to public transit in order to access congested areas like central business district (CBD). Driving to park-and-ride locations also acts as feeder service to light rail, commuter rail or other types of public transit. In order to make transit effective, it’s very important to locate the park-and-ride facilities at appropriate locations. So, finding optimal park and ride facility locations has become one of crucial tasks for planners as inappropriate locations may force people to drive and result in more congestion.

This study proposes a mixed integer nonlinear programming (MINLP) problem to find the optimal location of park-and-ride facilities in an urban network. The objective is to minimize the total system travel time while satisfying stochastic user equilibrium. The decision variables in this model are binary variables corresponding to candidate park-and-ride locations which take value 1 if the park-and-ride location is selected and 0 otherwise. Finally, the optimization problem is constrained by maximum number of park-and-ride locations to be built among candidate locations and also budget constraints.

To solve the given optimization problem a solution algorithm is proposed which consists of a stochastic traffic assignment to calculate the flow variables and a branch & bound algorithm to solve the relaxed integer programming problem. Finally, a hypothetical example is solved using the developed solution algorithm and results are presented.

Keywords: Park-and-ride, facility location, branch and bound, stochastic user equilibrium

Introduction  
Reducing congestion on roads is one of the primary concerns of traffic engineers and transportation planners. Researchers have proposed various solutions to this problem like prioritizing transit, increasing fuel taxes, levying tolls on congested highways and many more (1). It has been observed that certain areas like central business district (CBD) are more congested in comparison to others and park-and-ride facilities have been proved successful (19) in reducing congestion in those areas. People can use their private vehicle up to a certain stretch of the road, then park their vehicle at these facilities and switch to public transit in order to reach their destination. In this way, not only congestion, but also other costs such as travel time, environmental costs, etc. reduces.

The competence of park-and-ride facilities depends on their location in an urban network in order to draw more people to use them. The problem of finding the optimal park-and-ride locations is complicated (4), (2) as a lot of factors can affect these locations like position relative to CBD or primary activity centers, negative lot competition, travel characteristics to CBD, demographic factors, etc. Also, if they are not properly located, then these facilities may not be used and people may be forced to drive which will result in more congestion.

In this regard, various design guidelines have been developed based on the empirical studies, good planning and efficient operational strategies. For instance, (2), (3) and (4) give guidelines to decide the location of these facilities based on certain
rules like it should be located on upstream of congested areas, considering the geographical location of activity center served, auto versus transit cost, negative lot competition, transit connectivity, community integration, etc. However, these recommendations are sometimes perplexing and inconsistent among different researchers, for example, the distance of park-and-ride facility location from congestion varies as (6) suggests that park-and-ride facilities should be no closer than 5 to 6 km from the downtown areas whereas (4) recommend that park-and-ride facilities are adequately located at 10 miles away from central business district and similarly different recommendations are there in the design guidelines.

Hence, it requires some mathematical framework for solving the problem of finding the optimal locations of park-and-ride facilities in an urban network. In this regard, various mathematical models have been suggested in the literature. (7) evaluated the optimal location of park-and-ride facilities using profit maximization and social cost minimization. (8) considered three objective functions for locating park-and-ride facilities i.e. maximizing the demand coverage of the facility, minimizing the total cost between facility and major highways and maximizing the number of park-and-ride facilities while constraining total number of park-and-ride facility to be sited. (9) used a p-hub approach and considered every park-and-ride facility as a hub where people switch to transit from their cars in order to reach their destination. The aim was to maximize the park-and-ride usage while constraining the number of park and ride facilities to be built. Furthermore, some researchers have developed a bi-level optimization problem in order to consider more factors into the same problem. (13) suggested a bi-objective programming model for locating park-and-ride facilities which maximizes total flow using park-and-ride mode and minimizes the cost of constructing these facilities while keeping the constraints like allocating the demand to a particular park-and-ride facility and spatial equity constraints with respect to location of two park-and-ride facilities. Similarly, (15) developed a bi-level programming optimization model to site park-and-ride facilities. They maintained an upper optimization model for maximizing the social welfare (SW) which is sum of consumer surplus (CS) and producer surplus(PS) and used stochastic user equilibrium on the lower level.

The current research presents a novel approach to the problem which is a mixed integer non-linear programming (MINLP) problem. The objective is to minimize the total system travel time while satisfying the certain constraints like stochastic user equilibrium and budgetary constraints for building these facilities. The proposed methodology can be expanded as follows.

Methodology
Let $Y = \{y_n\}$ represent the set of candidate park-and-ride locations in a network where $y_n$ represents a binary decision variable which takes value 1 if park-and-ride is selected at location n and zero otherwise. Let $A$ represent the set of links $(i, j)$ and $N$ represent the set of nodes in the network. The following assumptions are made while formulating the optimization problem.

1. It is assumed that total demand from an origin to a destination is fixed. However, elastic demand can be considered for future research work.
2. The system consists of only two modes which is either to choose car only option or drive to a park-and-ride facility and switch to public transit in order to reach the destination. The share of transit trips with walking access is not impacted by the location of park-and-ride facilities because transit routes and stops remain the same.
3. After the park-and-ride location is known, modal split can be assumed to be constant (i.e. auto and park-and-ride demand is fixed). However, in the optimization problem, auto and park-and-ride demand is determined by the decision about park-and-ride locations.
4. As the park-and-ride location is changed, the traffic flow is changed due to users’ route choice behavior.
5. Transit travel time is fixed.

Objective Function
In order to make a transportation system efficient, total system travel time should be minimized. The objective of the current problem is therefore to minimize total system travel time while satisfying the stochastic user equilibrium and some other constraints. In this case, the total system travel time (TSTT) can be considered as the sum of two components which is total system travel time for auto trips and park-and-ride trips respectively:

$$TSTT = TSTT_{Auto} + TSTT_{P&R}$$

In order to achieve the expected maximum utility (17) while using the park-and-ride facilities, satisfaction function is given as log sum of exponential of the generalized cost function,
\[ cost(Y, h) = \frac{1}{\theta} \log e \left( \sum_n y_n e^{-\theta(Crn(k) + Cns)} \right), \quad \sum_n y_n \neq 0 \]

where \( h \) represents the vector of path flows. When the park-and-ride facility is located then \( y_n \) takes the value 1 and the exponential term becomes active. It is worth to note that the generalized cost consists of sum of travel time \( c_o \) from origin \( r \) to park and ride location \( n \) and travel time \( c_n \) from \( n \) to destination node \( s \). The dispersion or the scale parameter \( \theta \) has a value greater than 0. Also, we assume all the travel time functions to be differentiable and monotone. So, the total system travel time due to park-and-ride use can be calculated by multiplying the satisfaction function (Sheffi 1985) by demand.

\[ TSTT_{P&R} = \sum_{(r,s) \in Z^2} -\hat{d}^{rs} \log e \left( \sum_n y_n e^{-\theta(Crn(k) + Cns)} \right) \]

where \( \hat{d}^{rs} \) represents the park-and-ride demand rate from origin \( r \) to destination \( s \), \( h \) represents vector of path flows and \( Z \) is the set of zones. Moreover, the total system travel time due to automobile trips can be calculated by summing the product of demand and the path travel times over all the O-D pairs:

\[ TSTT_{Auto} = \sum_{(r,s) \in Z^2} d^{rs} c^{rs}(h) \]

where \( c^{rs} \) is the travel time from origin \( r \) to destination \( s \) using automobile and \( d^{rs} \) represents the auto demand from origin \( r \) to destination \( s \). Hence, our objective function for our optimization problem can be written as

\[ f(h, y, \theta) = \sum_{(r,s) \in Z^2} d^{rs} c^{rs}(h) + \sum_{(r,s) \in Z^2} -\hat{d}^{rs} \log e \left( \sum_n y_n e^{-\theta(Crn(k) + Cns)} \right) \]

**Constraints**

The given optimization model is subjected to several constraints. As we know due to budget constraints, we can only select a limited number of park-and-ride facilities among the candidate locations. Let the maximum number of park-and-ride facilities to be built be \( \bar{Y} \), therefore:

\[ \sum_n y_n \leq \bar{Y}, y_n \in Y \]

Before moving forward, it is supposed that OD trip demand is determined using Stochastic User Equilibrium (SUE). Multinomial Logit choice model can be utilized to calculate the auto demand \( d \) and park-and-ride demand \( \bar{d} \). For this, the probability of choosing any park-and-ride facility can be given as:

\[ P_{P&R}(h, y_n, \theta) = \frac{y_n e^{-\theta(Crn(k) + Cns)}}{\sum_n y_n e^{-\theta(Crn(k) + Cns)} + e^{-\theta C_{rs}}} \quad \forall (r, s) \in Z^2, \forall n \in Y \]

And similarly, the probability of choosing an auto route can be calculated as:

\[ P_{Auto}(h, y_n, \theta) = \frac{e^{-\theta C_{rs}(h)}}{\sum_n y_n e^{-\theta(Crn(k) + Cns)} + e^{-\theta C_{rs}}} \quad \forall (r, s) \in Z^2, \forall n \in Y \]

Note that the sum of all the probabilities \( P_{P&R} \) and \( P_{Auto} \) is equal to one. Hence demand \( d \) can be calculated as:

\[ d = \bar{d} P_{Auto}(h, Y, \theta) \]

Where \( \bar{d} \) is the total demand from an origin to a destination.

So, our optimization model for locating park-and-ride facilities can be summarized as below:

\[ \text{Min} \quad f(h, Y, \theta) = \sum_{(r,s) \in Z^2} d^{rs} c^{rs}(h) + \sum_{(r,s) \in Z^2} -\hat{d}^{rs} \log e \left( \sum_n y_n e^{-\theta(Crn(k) + Cns)} \right) \quad (1) \]

Subject to

\[ \sum_n y_n \leq \bar{Y}, y_n \in Y \]

\[ d = \bar{d} P_{Auto}(h, Y, \theta) \]

\[ h \geq 0 \]

\[ y_n = \{0, 1\} \quad \forall y_n \in Y \]
The following observations can be made for the proposed model.

1. In the given problem, the stochastic user equilibrium is a constraint to reflect the changes in the travel time by commuters’ path choice.
2. Given \( Y \), and assuming all the link travel time functions as monotonically increasing, continuously differentiable and separable in link flows, the given problem entails a unique SUE solution.
3. It is interesting to note that the relaxed problem (i.e. by fixing \( Y \)) is a convex optimization problem and is equivalent to stochastic assignment for park-and-ride lots.\(^{(21)}\)
4. Given all the assumptions, the SUE solution can be viewed as implicit function of \( Y \) and the given problem can be treated as:

\[
\begin{align*}
\text{Min } f(h^*, Y, \theta) \\
y_n = \{0, 1\} \forall n \in Y \\
\sum_n y_n \leq Y, n \in Y
\end{align*}
\]

which is an integer programming problem. The next section discussed a solution algorithm in detail.

Solution Algorithm
As we discussed the given minimization problem with implicit SUE constraints is an integer programming (IP) problem which is considered among the NP-hard problems. To solve this problem, various solution algorithms are suggested in the literature, one of the commonly used technique known as branch and bound (Land et. Al. 1960) can be used. For solving SUE, the method of successive averages (MSA) with stochastic network loading algorithm (Dial 1960) can be used. The solution algorithm for solving our proposed optimization model is discussed below:

1. Choose some initial integer feasible solution to the IP problem \( Y^0 \). Set counter \( c = 0 \) and go to step 2.
2. Solve the stochastic traffic assignment and calculate the flows \( h^* \).
   a. Set the iteration number \( k = 1 \). Calculate the link travel times by putting \( x_{ij}^k = 0 \) i.e. \( t_{ij}(0) \).
   b. Use the evaluated travel times in Dial’s algorithm and calculate the link flows \( x_{ij}^{k+1} \).
   c. Update the flow on links using \( x_{ij}^{k+1} = \frac{k-1}{k} x_{ij}^k + \frac{1}{k} x_{ij}^k \) and update the link travel times using \( t_{ij}(x_{ij}^{k+1}) \).
   d. Check the convergence using \( \frac{x_{ij}^{k+1} - x_{ij}^k}{x_{ij}^k} < \varepsilon \), stop, otherwise set \( k = k+1 \) and move to step 2(b).
3. Substitute the values of flows \( h^* \) in the original optimization problem (1) which produces a binary integer programming problem (2). Then Branch and Bound method can be used to solve for \( Y_c \). The algorithm steps are given below.
   a. In this method, we first solve the LP relaxation of binary IP by making all the binary variables continuous: \( 0 \leq y_k \leq 1, \forall k \).
   b. If an integer solution is found, then it is optimal to the IP, otherwise we can use the objective value at that solution as lower bound on the objective value. Then, we start branching the problem in two branches that is one by putting \( y_k = 0 \) and other by \( y_k = 1 \) and solve the problem again while keeping other variables relaxed as \( 0 \leq y_k \leq 1 \).
   c. At every step, the current solution is checked: if it is not better than the best solution found so far or there is not a feasible solution to the given branch, we prune that branch and do not go further into that.
   d. A gap is maintained to check out the convergence of solution at each step which is given below:
   \[
   \text{Integrality Gap} = \left| \frac{V^{IP} - V^{LP}}{V^{LP}} \right|
   \]
   where \( V^{IP} \) is the objective value of the integer solution and \( V^{LP} \) is the objective value based on the LP relaxation of the problem.
4. After calculating \( Y_c \), check the convergence: if \( Y_c = Y_{c-1} \), stop, otherwise set \( c=c+1 \) and go to step 2.

Numerical example
A small example is solved below using the solution algorithm for the proposed methodology. The test network (Fig 1) consists of five nodes \( N = \{1, 2, 3, 4, 5\} \) and seven links \( A = \{a, b, c, d, e, f, g\} \). Note that nodes 4 and 5 represent the park-and-ride candidate locations and links \( f \) and \( g \) represent transit links (shown in Fig. 1 by constant travel time function) in the network. Let \( Y = \{y_1, y_2\} \) be binary decision variables corresponding to each park-and-ride candidate location.
where \( y_1, y_2 \) take value 1 if park-and-ride is located, and 0 otherwise. The value of dispersion parameter \( \theta \) is assumed to be equal to 0.4. It is assumed that the total number of park-and-ride facilities to be built is equal to 1 i.e. \( \sum n y_n \leq 1 \).

The algorithm is initiated using initial basic feasible solution \( Y^0 = \{0, 1\} \). A python script was developed for calculating the flow variables using MSA with Dial's stochastic network loading algorithm. The results are given in Table1. Following Step 3 of solution algorithm by substituting SUE flow values in (1), produce a binary IP. This programming problem is solved using Branch and Bound method. For that CVXPY is used to solve the optimization problem which was done using another script developed for branch and bound method. For the current problem, the algorithm finds optimal solution as \( Y^1 = \{1, 0\} \) with an objective value of 59.4269. As the given optimal solution is not equal to the initial solution i.e. \( Y^1 = \{1, 0\} \) so we require further iterations. With new \( Y \) values, stochastic traffic assignment was solved again, results of which are given in Table1. Then a new IP is formulated using new results of flow values and iteration 2 of branch and bound method is followed using the same steps. This time \( Y^2 = \{1, 0\} \) which is equal to \( Y^1 = \{1, 0\} \), that means we can stop here. Hence, we need to build park-and-ride facility at location 1.

<table>
<thead>
<tr>
<th>Link Flows</th>
<th>Y</th>
<th>Objective Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration</td>
<td>xa(12)</td>
<td>xd</td>
</tr>
<tr>
<td>1</td>
<td>3.13917</td>
<td>1.612246</td>
</tr>
<tr>
<td>2</td>
<td>3.395884</td>
<td>0</td>
</tr>
</tbody>
</table>

Conclusions and Recommendations
This research formulates the problem of finding the optimal subset of park-and-ride locations as an optimization problem with SUE constraints. The objective function of the model captures the maximum utility of available park-and-ride locations and minimizes the total vehicle-hours users spend in the system. A solution algorithm is developed for solving the given problem based on branch and bound, where SUE constraints are considered implicit. Finally, a small example is solved for illustration purposes. The algorithm was found to be effective in solving the problem. Followings are the recommendations for future research.

1. The problem was solved for a hypothetical network and needs to be tested on a real network.
2. Travel cost function consists of only travel time component. It is recommended to use more generalized cost function by including other costs such as parking costs, etc.
3. The solution algorithm can be improved and its computational efficiency is not evaluated yet.
References


15. Sheffi, Yosef 1985, Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods


17. The effectiveness and sustainability of park and ride. Histrionics Publication


Reliable Facility Location Design with Imperfect Information: Continuum and Discrete Models

Lifen Yun, Hongqiang Fan, Xiaopeng Li

Civil and Environmental Engineering, University of South Florida, Tampa, 33620, USA,

In the reliable facility location design, most of studies assume that a customer can access the perfect information of the functioning state of every facility in real time and thus always knows the most appropriate facility to visit (e.g., this customer’s closest functioning facility). However, this assumption may not always reflect the reality due to customer unpreparedness, technology limitations, and institutional barriers. Further, accessing real-time facility states becomes even more difficult under unexpected emergency scenarios due to likely communication and infrastructure disruptions. Therefore, considering the imperfect information in the study of reliable facility location design is necessary to balance the trade-off between one-time investment and long-term expected operation cost.

Under imperfect information, i.e., when real-time facility states are unknown, a customer may have to try a series of pre-specified facilities (which may or may not be functioning) until finding a functioning one to obtain the service or giving up on trying to receive a penalty instead. In realistic world, customer always returned her initial location when she has obtained the service and given up looking for the service. Therefore the customer’s trip for the service is always a two-way trip which includes the departure trip and the return trip. In the perfect information context, the departure trip and return trip are simply identical. Thus, the return trip is not necessary in the transportation cost formulation. But in the imperfect information context, they are totally different and the return trip need be considered in the transportation cost formulation. Therefore, this paper proposes two models, which are extended the discrete model previously developed by the authors, for the reliable facility location design with imperfect information and two-way trip.

We first propose a scalable continuum approximation (CA) model for a reliable facility location design problem under imperfect information considering two-way trips. In the heterogeneous plane, the CA model is hard to be solve directly. First, we decompose the very complex origin location problem into a set of relatively simple sub-problems with spatially homogenous settings.

In the initial homogenous plane (IHP), we assume the initial service areas \( A \) (i.e., the area a facility serves when all facilities are functioning) of facilities form a regular hexagonal tessellation. Each facility is located at the center of a regular hexagon. The hexagon tessellation is proven to be the optimal facility layout for traditional deterministic facility location problems when the transportation cost is proportional to the Euclidean distance. Therefore, we can easily obtain the unit-area facility opening cost and the unit-area penalty cost.

\[
C^F = f/A \tag{1}
\]

\[1\] Corresponding author (Email: xiaopengli@usf.edu; Tel: 1-813-974-0778)
\[ C^p = \lambda \varphi q^{R+1} \]  

For the unit-area transportation cost, it is decided by the customers' visiting sequences. Different from the IHP analysis under perfect information, the challenge of our proposed CA model is that customers in different parts of the central service area may have different visiting sequences. To overcome this challenge, we divide one hexagon as the central hexagon in IHP into the twelve identical sectors. Note that since these sectors are symmetric around the central axis, the costs associated with these sectors are identical. Therefore, we choose the upper-right sector as the analysis area for calculating the unit-area transportation cost as shown in the following figure.

While it is not intuitive to exactly determine each customer's optimal visiting sequence (OVS), we can construct a near-optimum visiting sequence (NOVS) for all customers in the upper-right sector that we later show yields a near optimum solution. We index facilities (or hexagons) around the central facility. The central facility is indexed as 0, the facility immediately above the central one is indexed as 1, and then the indexes increase clockwise along the spiral traversing all these facilities sequentially. The NOVS assumes that the order for any customers in the analysis area to visit these facilities is consistent with these indexes. We can get a near-optimal solution according to the NOVS. Due to the optimal solution is hard to obtain, we construct a lower bound which is obtained easily to the optimal solution. We prove that the gap between the near-optimal solution and the lower bound is small. Therefore, we deem the near-optimal solution is near the optimal solution and can be used to approximate the optimal solution in the unit-area transportation formulation.

\[ C^T = C^{T-NO} = \lambda A^{1/2} \left( \beta_0 + R \left( \frac{4}{3} \right)^{1/4} + \beta_r \right) q^{R+1} + \sum_{r=0}^{R} \left( \beta_0 + r \left( \frac{4}{3} \right)^{1/4} + \beta_r \right) q^r (1-q) \]  

Finally, we get the unit-area system cost function for the IHP problem.

\[ C(A) = C^f + C^p + C^T = f/A + \lambda \varphi q^{R+1} + \lambda A^{1/2} \left( \beta_0 + R \left( \frac{4}{3} \right)^{1/4} + \beta_r \right) q^{R+1} + \sum_{r=0}^{R} \left( \beta_0 + r \left( \frac{4}{3} \right)^{1/4} + \beta_r \right) q^r (1-q) \]  

After we solve the IHP problem, we discuss how to apply the IHP results to the original finite heterogeneous plane. We assume that all the parameters vary relatively slowly in the
heterogeneous space. Instead of looking for the discrete location design scheme, we look for a continuous function that approximates the initial service area size of a facility near. Since we assume that the facilities are densely located, the size of the heterogeneous plane should be much larger than the service area of any facility. Thus the heterogeneity around the boundary of the heterogeneous plane can be ignored without affecting much the total cost. When the parameters are approximately constant over a region comparable to the size of several influence areas, the influence area size should also be approximately constant on that scale as well. In this case, we apply the function of the unit-area system cost for the IHP problem to each neighborhood of an arbitrary location (i.e., imagining that this neighborhood is part of an IHP). Given the parameters of the original heterogeneous plane, we obtain the system cost function of serving a unit area near an arbitrary location.

\[
C(x, A(x)) := f(x)/A(x) + \lambda(x) \varphi(x) q^{R+1}(x) + \lambda(x) A^{1/2}(x)
\]

\[
\left( \beta_0 + R \left( \frac{4}{3} \right)^{1/4} + \beta_r \right) q^{R+1}(x) + \sum_{r=1}^{\infty} \left( \beta_0 + r \left( \frac{4}{3} \right)^{1/4} + \beta_r \right) q^r(x)(1-q(x)) \quad (5)
\]

We can get the optimal service area \( A^*(x) \) for an arbitrary location by solving the optimal problem of \( C(x, A(x)) \). Then the optimal system cost in original finite heterogeneous plane can be approximated by integrating the system cost function of serving a unit area near an arbitrary location.

\[
C^* = \int_{x \in S} C(x, A^*(x)) dx \quad (6)
\]

The total number of optimal facilities can be estimated by the inverse of the optimal service area.

\[
N^* \approx \int_{x \in S} \left[ A^*(x) \right]^{-1} dx \quad (7)
\]

A sweeping method which requires only a linear computational time and space is used to obtain the discrete location solution for the CA model.

For comparison, we also propose a counterpart discrete model to the proposed CA model. Before constructing discrete model, we convert the continuous parameters into the discrete parameters. In the process of constructing discrete model, we first formulate the total system cost function with imperfect information and a two-way trip under a given facility location design scheme. Then, we introduce two sets of auxiliary decision variables to specify facility assignments in our proposed discrete model. We also need introduce a set of probability variables to specify visiting probabilities because each facility has its own disruption probability in the heterogeneous plane. With these methods, we obtain a nonlinear discrete model for the reliable facility location design problem.

\[
\min_{x_i, y_{ij}} \sum_{j \in J} f_j y_j + \sum_{i \in I} \lambda_i p_{ij} \varphi_i + \sum_{i \in I} \sum_{j \in J} \sum_{k \in I} \left( \sum_{r=1}^{\infty} \left( d_{ij} + (1-q) d_{ij} \right) p_{iij} x_{ijr} + \sum_{r=1}^{\infty} \left( d_{ij} p_{iij} x_{ijr} \right) \right) \quad (8)
\]
subject to

\[ x_{ij} + \sum_{j' \in J_j} \sum_{r=1}^{R} x_{ij' r} \leq y_j, \forall i \in I, j \in J \]  

(9)

\[ \sum_{j \in J_i} x_{ij} = 1, \forall i \in I \]

(10)

\[ x_{ij} = \sum_{j' \in J_j} x_{ij'1}, \forall i \in I, j \in J \]

(11)

\[ \sum_{j' \in J_j} x_{ij'j(r-1)} = \sum_{j' \in J_j} x_{ij' r}, \forall i \in I, j \in J, j' \in J \setminus j, r = 2,3,\ldots, R \]

(12)

\[ \sum_{j \in J} x_{ijh} = 1, \forall i \in I \]

(13)

\[ p_{ij'1} = q_j, \forall i \in I, j \in J, j' \in J \setminus j \]

(14)

\[ p_{ij'r} = q_j \sum_{j' \in J_j} p_{ij'j(r-1)} x_{ij'j(r-1)}, \forall i \in I, j \in J, j' \in J \setminus j, r = 2,3,\ldots, R \]

(15)

\[ y_j \in \{0,1\}, \forall j \in J \]

(16)

\[ x_{ij} \in \{0,1\}, \forall i \in I, j \in J \]

(17)

\[ x_{ij' r} \in \{0,1\}, \forall i \in I, j \in J, j' \in J \setminus j, r = 1,2,\ldots, R \]

(18)

For solving the discrete model easily, we apply the linearization technique to change the nonlinear discrete model to the linear discrete model. We replace each \( p_{ij', r} \) with new variable \( w_{ij', r} \). A set of new constraints showed in the follow is added to the model (8)-(18) to enforce \( w_{ij', r} := p_{ij', r} x_{ij' r}, \forall i \in I, j \in J, j' \in J \setminus j, r = 1,2,\ldots, R \).

\[ w_{ij' r} \leq p_{ij', r}, \forall i \in I, j \in J, j' \in J \setminus j, r = 1,2,\ldots, R \]

(19)

\[ w_{ij' r} \leq x_{ij' r}, \forall i \in I, j \in J, j' \in J \setminus j, r = 1,2,\ldots, R \]

(20)

\[ w_{ij' r} \geq 0, \forall i \in I, j \in J, j' \in J \setminus j, r = 1,2,\ldots, R \]

(21)

\[ w_{ij' r} \geq p_{ij', r} + x_{ij' r} - 1, \forall i \in I, j \in J, j' \in J \setminus j, r = 1,2,\ldots, R \]

(22)

The linearize formulation of model (8)-(18) is stated below:

\[
\min_{x, x', y} \sum_{i \in I} \sum_{j \in J} f_{ij} y_j + \sum_{i \in I} \lambda_i w_{ij'} \varphi_i + \sum_{i \in I} \lambda_i \sum_{j \in J} \left( \sum_{j' \in J \setminus j} \left( \sum_{r=1}^{R} \left( d_{ij'} + (1-q)d_{ij'} w_{ij' r} \right) \right) \right) \]

(23)
Subject to (9)-(14), (16)-(22) and
\[ p_{ijr} = q_j \sum_{j' \neq j} w_{ij'j(r-1)}, \forall i \in I, j \in J, j' \in J \setminus j, r = 2,3,\ldots,R \]  
\hspace{1cm} (24)

A customized Lagrangian relaxation algorithm are proposed to solve the proposed discrete model by relaxing the constrains (9).

Numerical studies are conducted to test algorithm efficiency and solution quality and discuss managerial insights. Over all, this proposed models both can solve the studied reliable location problem of reasonable sizes, while the CA model has superior efficiency and scalability compared with the discrete model. Nonetheless, we would like to note that the CA solution, though close to the true optimum in many practical problem instances, is not guaranteed to be always optimal, and the discrete counterpart model is still needed for the exact optimal solution or a rigorous optimality bound. Further, it reveals the managerial insights into an approximated problem with homogeneous settings more intuitively (e.g., how a user chooses to visit facilities in a homogeneous plane). The influence of imperfect information and return trip on the total system cost are investigated deeply by the proposed CA model. Ignoring these two conditions may underestimate the total system cost. The feature of visiting sequence is also investigated. It is very similar to our realistic situations. If the customer needs return her initial location, she will look for the service around her initial location. If she does not return her initial location, she may find the service far away the initial location.

(a) The visiting sequence with inbound travel  
(b) The visiting sequence without inbound travel

The results of sensitivity analyses indicate that the optimal location design under imperfect information and two-way trip shall vary significantly as the disruption probability considerably increases. The other result shows that the benefit of an additional backup facility to a customer generally decreases exponentially as the assignment level increases, since his no service probability is an exponentially decreasing function of his total assigned facilities. These results indicates that the proposed CA model has a robust performance for large-scale reliable facility location design problems.
Location-Routing Problems with Economies of Scale

1. Introduction

The location-routing problem (LRP), an important application is supply chain management, combines the facility location problem (decisions on the siting of one or more facilities) and the execution of vehicle routes to customers from the chosen, open facilities. Since each facility will act as a depot for particular routes that are operated, the transportation impact of the LRP will thus be better accounted for, when facility choices are made in this light. The set of feasible sites is an input to the model; the locations of customers are assumed known and fixed.

1.1 Location-Routing Problem

The standard, or single-echelon, LRP involves routing at only one level. The two-echelon location-routing model (Prodhon and Prins, 2014; Drexl and Schneider, 2015) includes factory, DCs, and customers, with consideration of routing at each echelon (between factory and DCs, and between DCs and customers). Most references develop routes only to customers, from depots (warehouses/DCs). Many papers in fact do not worry about the flow of goods into the DCs. Rather, that supply of product is given, and routes are constructed from each DC to a number of customers (retailers). The routing is only at that level, hence it is a single-echelon location-routing model that is studied.

A few papers do consider a factory or supply source for the distribution centers. In both Perl and Daskin (1985) and Bookbinder and Reece (1988), there are multiple supply sources for DCs; products are shipped directly from factories to each DC. Routes are then designed between DCs and customers. The majority of published papers concern the capacitated LRP. In our numerical studies, all depots and vehicles have a finite capacity.

Nagy and Salhi (2007) summarize the attributes of different location-routing models. Borges Lopes et al. (2013) classify each LRP model in the form of a taxonomy. Research on location-routing problems for the years 2007-2013 is surveyed by Prodhon and Prins (2014). Drexl and Schneider (2015), in the most recent survey on the location-routing problem, discuss both the standard LRP (single-echelon) and two-echelon LRPs.

1.2 Economies of Scale

Scale economies, i.e. concave costs, have been studied in the context of problems such as warehouse or plant operations. To the best of our knowledge, only Melechovsky et al. (2005) has previously considered economies of scale in dealing with the location-routing problem. With routing at only one level, they propose a non-linear variable cost function for the throughput of products handled in each facility (DC), incorporating the economies of scale in that single-
The result is a mixed-integer model for which they propose a metaheuristic method that combines variable neighborhood search with Tabu search.

Naturally, in practice, cost is incurred when goods move in or out of a facility. However, this facility-throughput cost is only sometimes included in the LRP literature. And except in Melechovsky et al. (2005) and the present paper, that cost has not considered economies of scale. Most often, that variable cost is taken to be a linear function of throughput at the DC.

To represent the economies of scale that are present in many real-world facilities, the operating cost, i.e. the variable cost of operation of each DC, should be a concave function of facility throughput. Our research thus focuses on the influence of those economies of scale at a distribution center, when vehicle routes must also be considered. Throughout, the terms, “variable cost” and “operating cost,” shall be synonymous.

1.3 Methods for Solving Location-Routing Problems

Only a small number of publications employ exact solution methods for the LRP; both the location problem and the vehicle routing problem are of course NP-hard. Prodhon and Prins (2014) note that lower bounds for the LRP have been obtained by just a few researchers, e.g. Belanguer et al. (2011) and Contardo et al. (2014), namely those implementing exact methodologies. Most location-routing problems therefore employ efficient metaheuristics.

Wu et al. (2002) solve the multi-depot LRP, with limited homogeneous fleets, by a simulated-annealing approach. Prins et al. (2006a) design a greedy randomized adaptive search procedure for a general LRP with uncapacitated vehicles, based on the Clarke-Wright algorithm combined with a learning process. For capacitated depots and vehicles, Prins et al. (2007) solve location-routing problems on a complete network by a heuristic decomposition method. To minimize the total fixed costs of depots plus variable costs of routes, the latter authors alternate between a facility-location phase, solved by Lagrangian relaxation, and a routing phase, handled by a granular Tabu search. Prins et al. (2007) find that their method can improve 80% of the values obtained by others on three publicly available data sets, two of which are utilized in our work.

2 Results

Following experiments with metaheuristic methods, we employed a genetic algorithm with ant colony optimization. Tables 1 and 2 present our numerical findings for the standard data sets of Barreto (2004) and Prins et al. (2006a), respectively. Naturally, we first had to enhance those instances by including variable costs, so that we could study the influence of economies of scale. These economies, when present (model [LRPES]), are modelled by a concave power function $f_i(x) = \gamma x_i^\delta$, where $\delta = 2/3$ and $x_i$ is the total throughout of facility $i$. Without economies of
scale, in model \([LRPV]\), the variable cost of a DC is expressed as a unit operating cost multiplied by the total demand (throughput) served. Note that although both of our mathematical models include facility operating costs, only the second has scale economies. Both formulations have the objective of achieving minimal total cost. Facilities can have different capacities, but are the same in all other aspects. In other words, each facility has the same operating-cost function.

The two models are thus very similar; differences lie in the objective functions. The objective function of each model is the sum of the operating cost including fixed cost, variable cost of operation, and transportation cost between DCs and customers. The variable cost in \([LRPV]\) is calculated by multiplying the total throughput by the (constant) unit operating cost. The total operating cost in \([LRPV]\) can thus be calculated directly by the total demand of all customers. However, the model \([LRPES]\) employs the power function, and hence has an additional decision variable: The marginal operating cost of an opened DC changes with the number of units of demand satisfied by that DC.

Otherwise, the two models satisfy the same constraints: Each customer will be connected to a unique predecessor by a single vehicle; the capacity of each vehicle will be respected. Two relations ensure the continuity of routes, and that all vehicles travel back to the origin, i.e. the DC. Sub-tours are eliminated in the usual way. Only those locations selected as DCs can become the origin of a route; the sum of customers’ demands satisfied by a given DC cannot exceed the capacity of that DC. The final two constraints show the binary nature of decision variables.

Computational results are given in Tables 1 and 2, in which \(n\) is the number of customers and \(m\) is the number of possible sites for DCs. We have generally found that, in the presence of scale economies, fewer facilities are opened. Although that causes an increase in the DC \(\rightarrow\) customer routing costs, the total costs are diminished relative to when there are no economies of scale.

We remark that, for most of the instances \textit{without variable costs} that we solve, our solution is typically within 3\% of the best known result for the respective instance.
<table>
<thead>
<tr>
<th>Instance</th>
<th>n</th>
<th>m</th>
<th>Number of DC opened</th>
<th>Operating Cost</th>
<th>Total Cost</th>
<th>CPU(s)</th>
<th>Number of DC opened</th>
<th>Operating Cost</th>
<th>Total Cost</th>
<th>CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christofides69 50</td>
<td>5</td>
<td>4</td>
<td>1166</td>
<td>1790</td>
<td>75.24</td>
<td></td>
<td>2</td>
<td>1064</td>
<td>1650</td>
<td>78.81</td>
</tr>
<tr>
<td>Christofides69 75</td>
<td>10</td>
<td>4</td>
<td>2046</td>
<td>2926</td>
<td>130.52</td>
<td></td>
<td>2</td>
<td>1549</td>
<td>2391</td>
<td>133.28</td>
</tr>
<tr>
<td>Christofides69 100</td>
<td>10</td>
<td>4</td>
<td>2187</td>
<td>3079</td>
<td>342.57</td>
<td></td>
<td>2</td>
<td>1625</td>
<td>2479</td>
<td>19.99</td>
</tr>
<tr>
<td>Daskin95     88</td>
<td>8</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Daskin95     150</td>
<td>10</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gaskell67    21</td>
<td>5</td>
<td>2</td>
<td>1350</td>
<td>1769</td>
<td>26.37</td>
<td></td>
<td>2</td>
<td>1034</td>
<td>1428</td>
<td>26.92</td>
</tr>
<tr>
<td>Gaskell67    22</td>
<td>5</td>
<td>3</td>
<td>611</td>
<td>1259</td>
<td>32.04</td>
<td></td>
<td>2</td>
<td>594</td>
<td>1200</td>
<td>33.48</td>
</tr>
<tr>
<td>Gaskell67    29</td>
<td>5</td>
<td>3</td>
<td>765</td>
<td>1318</td>
<td>35.68</td>
<td></td>
<td>2</td>
<td>687</td>
<td>1202</td>
<td>36.33</td>
</tr>
<tr>
<td>Gaskell67    32</td>
<td>5</td>
<td>2</td>
<td>1762</td>
<td>2332</td>
<td>33.37</td>
<td></td>
<td>2</td>
<td>1219</td>
<td>1797</td>
<td>33.60</td>
</tr>
<tr>
<td>Gaskell67    32</td>
<td>5</td>
<td>2</td>
<td>1762</td>
<td>2296</td>
<td>65.03</td>
<td></td>
<td>1</td>
<td>952</td>
<td>1460</td>
<td>67.42</td>
</tr>
<tr>
<td>Gaskell67    36</td>
<td>5</td>
<td>2</td>
<td>1350</td>
<td>1832</td>
<td>70.36</td>
<td></td>
<td>1</td>
<td>935</td>
<td>1440</td>
<td>69.85</td>
</tr>
<tr>
<td>Min92       27</td>
<td>5</td>
<td>2</td>
<td>12615</td>
<td>15675</td>
<td>60.90</td>
<td></td>
<td>2</td>
<td>5211</td>
<td>8297</td>
<td>62.98</td>
</tr>
<tr>
<td>Min92       134</td>
<td>8</td>
<td>4</td>
<td>11866</td>
<td>17825</td>
<td>565.21</td>
<td></td>
<td>3</td>
<td>5729</td>
<td>11550</td>
<td>569.02</td>
</tr>
</tbody>
</table>

Tab. 1: Results for the instances of Barreto (2004)-with variable cost
<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>W</th>
<th>DC opened</th>
<th>Operating Cost</th>
<th>Total Cost</th>
<th>CPU(s)</th>
<th>DC opened</th>
<th>Operating Cost</th>
<th>Total Cost</th>
<th>CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>5</td>
<td>70</td>
<td>3</td>
<td>3,780</td>
<td>58,573</td>
<td>19.27</td>
<td>3</td>
<td>3,672</td>
<td>58,465</td>
<td>19.64</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>150</td>
<td>2</td>
<td>3,696</td>
<td>42,800</td>
<td>19.55</td>
<td>3</td>
<td>3,173</td>
<td>42,368</td>
<td>20.09</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>70</td>
<td>3</td>
<td>3,720</td>
<td>52,561</td>
<td>18.98</td>
<td>3</td>
<td>3,675</td>
<td>52,680</td>
<td>19.99</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>150</td>
<td>2</td>
<td>3,624</td>
<td>41,166</td>
<td>19.74</td>
<td>3</td>
<td>3,122</td>
<td>42,667</td>
<td>20.86</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>70</td>
<td>3</td>
<td>9,072</td>
<td>99,684</td>
<td>36.14</td>
<td>3</td>
<td>6,585</td>
<td>98,082</td>
<td>38.22</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>150</td>
<td>3</td>
<td>9,072</td>
<td>79,572</td>
<td>33.99</td>
<td>2</td>
<td>5,756</td>
<td>72,275</td>
<td>36.13</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>70</td>
<td>3</td>
<td>9,300</td>
<td>98,286</td>
<td>31.40</td>
<td>3</td>
<td>6,701</td>
<td>97,012</td>
<td>32.21</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>150</td>
<td>3</td>
<td>9,300</td>
<td>78,042</td>
<td>30.88</td>
<td>3</td>
<td>6,714</td>
<td>75,831</td>
<td>32.55</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>70</td>
<td>3</td>
<td>9,228</td>
<td>93,880</td>
<td>35.12</td>
<td>3</td>
<td>6,656</td>
<td>91,737</td>
<td>37.60</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>150</td>
<td>3</td>
<td>9,396</td>
<td>62,175</td>
<td>38.22</td>
<td>3</td>
<td>6,736</td>
<td>58,714</td>
<td>39.59</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>70</td>
<td>2</td>
<td>9,132</td>
<td>96,532</td>
<td>40.37</td>
<td>2</td>
<td>5,775</td>
<td>92,988</td>
<td>43.04</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>150</td>
<td>2</td>
<td>9,132</td>
<td>71,169</td>
<td>37.90</td>
<td>2</td>
<td>5,770</td>
<td>68,270</td>
<td>39.88</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>70</td>
<td>3</td>
<td>18,996</td>
<td>298,096</td>
<td>210.02</td>
<td>3</td>
<td>10,778</td>
<td>291,056</td>
<td>214.79</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>150</td>
<td>3</td>
<td>18,996</td>
<td>235,668</td>
<td>201.88</td>
<td>3</td>
<td>10,786</td>
<td>227,876</td>
<td>206.08</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>70</td>
<td>2</td>
<td>18,696</td>
<td>218,006</td>
<td>219.53</td>
<td>2</td>
<td>9,312</td>
<td>209,196</td>
<td>223.21</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>150</td>
<td>2</td>
<td>18,696</td>
<td>177,896</td>
<td>192.04</td>
<td>2</td>
<td>9,331</td>
<td>168,686</td>
<td>196.75</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>70</td>
<td>2</td>
<td>18,744</td>
<td>222,190</td>
<td>213.77</td>
<td>2</td>
<td>9,350</td>
<td>212,937</td>
<td>216.68</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>150</td>
<td>2</td>
<td>18,744</td>
<td>173,270</td>
<td>195.48</td>
<td>2</td>
<td>9,323</td>
<td>164,632</td>
<td>198.39</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>70</td>
<td>3</td>
<td>19,320</td>
<td>311,209</td>
<td>190.64</td>
<td>3</td>
<td>10,929</td>
<td>310,816</td>
<td>192.63</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>150</td>
<td>3</td>
<td>19,320</td>
<td>255,322</td>
<td>178.23</td>
<td>3</td>
<td>10,896</td>
<td>247,428</td>
<td>183.52</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>70</td>
<td>3</td>
<td>18,432</td>
<td>264,950</td>
<td>188.10</td>
<td>3</td>
<td>10,559</td>
<td>264,615</td>
<td>190.38</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>150</td>
<td>3</td>
<td>18,432</td>
<td>222,964</td>
<td>169.21</td>
<td>3</td>
<td>10,598</td>
<td>220,593</td>
<td>178.06</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>70</td>
<td>3</td>
<td>18,480</td>
<td>278,280</td>
<td>201.16</td>
<td>3</td>
<td>10,590</td>
<td>281,589</td>
<td>202.95</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>150</td>
<td>3</td>
<td>18,480</td>
<td>220,653</td>
<td>180.85</td>
<td>3</td>
<td>10,571</td>
<td>225,884</td>
<td>185.78</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>70</td>
<td>4</td>
<td>37,176</td>
<td>575,102</td>
<td>1254.23</td>
<td>3</td>
<td>16,869</td>
<td>508,779</td>
<td>1263.47</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>150</td>
<td>4</td>
<td>37,176</td>
<td>479,962</td>
<td>1184.43</td>
<td>3</td>
<td>16,848</td>
<td>432,631</td>
<td>1121.55</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>70</td>
<td>3</td>
<td>37,212</td>
<td>500,163</td>
<td>1215.40</td>
<td>3</td>
<td>16,845</td>
<td>508,816</td>
<td>1219.09</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>150</td>
<td>3</td>
<td>37,212</td>
<td>415,182</td>
<td>1014.76</td>
<td>3</td>
<td>16,872</td>
<td>395,152</td>
<td>1020.30</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>70</td>
<td>4</td>
<td>36,924</td>
<td>584,012</td>
<td>1219.65</td>
<td>3</td>
<td>16,785</td>
<td>516,815</td>
<td>1222.75</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>150</td>
<td>4</td>
<td>36,924</td>
<td>470,918</td>
<td>1040.01</td>
<td>3</td>
<td>16,770</td>
<td>392,745</td>
<td>1050.98</td>
</tr>
</tbody>
</table>

Tab. 2: Results for instances of Prins et al. (2006a), with addition of variable cost
Exact and approximate optimal route set generation in sensor location problems

Marco Rinaldi, Ph.D., Francesco Viti, Associate Professor (corresponding author)
University of Luxembourg, Research Unit of Engineering Science, Luxembourg (Luxembourg)

Background
The main goal of Network Sensor Location Problems (NSLP) is to determine the minimum set of observed link, route or OD flows that can be measured to provide information on the remaining non-observed link (or route, or OD) flows. Solutions to these problems are approached in the literature by exploiting the fundamental relationships between the three sets of variables (link flows, route flows, OD flows), derived from conservation of vehicles principles. In this work we focus on the link flow inference problem, i.e. identifying a (smallest) set of independent links able to fully determine the flow on other links in the network (see Castillo et al., 2015 for a detailed overview).

Existing approaches to compute link flow inference solutions are subdivided in methods that exploit node-link relations (Ng, 2012) or link-route relations (Castillo et al., 2008; Hu et al., 2009). This latter category of approaches has been shown to potentially identify more efficient solutions. The work of (Viti et al., 2014) showed that, for mid- and large-sized networks, node-based approaches tend to recommend a systematically higher number of sensors to install if compared to their route-based counterparts. A fundamental reason for this systematic difference lies in the richer information offered by link-route relations if compared to link-node relations. In (Castillo et al., 2014) the authors found that the minimum number of linearly independent links to be observed can be further reduced if a set of linearly independent paths is used. More recently, in (Rinaldi et al., 2015) we refined these results by empirically assessing the influence that route generation policies have on the shape and structure of the full observability solution. We specifically concentrated on studying how the amount and quality of information in either full or partial information solutions depends on the size and the composition of the route set. The main finding of the paper was that the overall information content tends to increase as the route set is expanded, but the rate of growth reduces non-linearly with the route set size, suggesting that there is an upper bound above which no new information is gained by including additional routes in the route set. Moreover, enumerating routes according to algebraic independence principles resulted in overall better-informative observability solutions with respect to standard enumeration techniques, a conclusion entirely in line with the findings of (Castillo et al., 2014).

The main contribution of this paper is to provide theoretical foundations to our previous empirical observations. In particular, we formulate an optimisation problem seeking to determine the maximum independent route set for a given network. Further, we show that, by solving the novel optimisation problem, we are able to systematically find full observability solutions exhibiting very high information content while requiring as much as or fewer sensors than other existing approaches. We show the effectiveness of the methodology through small toy examples.

Methodology

Link flow inference problems exploit the link-route flow relation (see e.g. (Cascetta, 2009)):

\[ \mathbf{v} = \mathbf{A} \mathbf{h} \tag{1} \]

with \( \mathbf{v} \) the link flow vector, \( \mathbf{h} \) the route flow vector. Link-route relationships are represented in form of incidence matrices \( \mathbf{A} \in \mathbb{R}^{\left| R \right| \times \left| \mathcal{R} \right|} \), where the matrix’s columns represent the \( \left| \mathcal{R} \right| \) routes composing the route set \( \mathcal{R} \). One approach to solve such problems is to perform opportune matrix manipulations, i.e. by swapping or sorting rows or columns of \( \mathbf{A} \), while carefully maintaining the original relations between the variables as in the original set of relations (1). This approach (which we will refer to as pivoting procedure in this paper) has been proposed originally in (Castillo et al., 2001) for power networks.

In order to understand how different route sets \( \mathcal{R} \) influence observability problems, in this approach we characterize the information brought in by each new route as pertaining to one of three sets:

- Non Redundant (NR) set: a route \( r_i \in \mathcal{R} \) pertains to the NR set if at least one of its links was not previously included in the route set (i.e. the route includes new, previously unavailable direct measurement information);
- Redundant while Informative (RI) set: a route \( r_i \in \mathcal{R} \) pertains to the RI set if its inclusion in the route set allows deriving algebraic interactions between previously non-correlated links (i.e. the route includes new previously unavailable indirect relationship information);
- Purely Redundant (PR) set: a route \( r_i \in \mathcal{R} \) pertains to the PR set if its inclusion in the route set \( \mathcal{R} \) brings no further information to the system.
In Figure 1 we showcase three simple examples of route sets for a small network: Fig. 1(a) shows two routes both pertaining to the NR set, in Fig. 1(b) we show how a different choice for one of these two routes yields instead RI information, while in Fig. 1(c) we show how a poorly chosen third route would consist of PR information. In general, an arbitrarily generated route set (including the full route set) will be a union of all three subsets:

\[ R = \{r_1, r_2, \ldots, r_n\} = [NR \cup RI \cup PR] \]  

(2)

Figure 1: example of route set with (a) non-redundant information, (b) Non Redundant and Redundant while Informative, and (c) Non Redundant, Redundant while Informative and Purely Redundant.

Determining the minimum independent route set yields a solution where \( A = [NR] \) and \( [RI \cup PR] = \emptyset \). Our intuition is that when extending this initial route set with further independent routes, both NR and RI sets will increase in size, but, if the independence constraint is observed, the PR set will still remain empty. Specifically, this condition is verified if and only if all routes (and combinations thereof) remain independent from one another.

We form the following hypothesis: “determining the maximum independent route set for a given network yields an optimal balance of NR and RI information while minimizing PR information”. In order to validate said hypothesis, we develop a methodology capable of constructing a set of routes \( R \) such that the three following conditions are verified:

i. The set is maximal in terms of cardinality
ii. All routes are independent from one another
iii. All combinations of routes are independent from one another

Given a generic transportation network \( \mathcal{N} \) described by a directed graph \( \mathcal{N} \sim G(L,V) \) with \( L \) the set of links and \( V \) the set of nodes and \( R = \{r_1, r_2, \ldots, r_n\} \) the set of routes, we define independence between routes as follows:

\[
\begin{align*}
  r_i &= \{l_{i1}, l_{i2}, \ldots, l_{in}\} \\
  r_j &= \{l_{j1}, l_{j2}, \ldots, l_{jn}\} \\
  \text{INDEP}(r_i, r_j) &\iff \exists l_k : l_k \in r_i \land l_k \notin r_j \\
  \text{INDEP}(r_i, r_j) &\iff \text{INDEP}(r_j, r_i)
\end{align*}
\]

(3)

given a route \( r_i \) composed by a set of links \( n \) and a route \( r_j \) composed by a set of links \( z \) (where \( n, z \in L \) and \( |n| \leq |L| \) and \( |z| \leq |L| \)), the two routes \( r_i \) and \( r_j \) are independent from one another if and only if there exists a link \( l_k \) which is included in route \( r_i \) and not included in route \( r_j \).

We formulate the following optimization problem:
max | \( R \) | \\
\text{s.t.} \\
\begin{align*}
\text{INDEP}(r_i, r_j) \forall i, j \in R & \\
\text{INDEP}([r_i \cup r_j], r_k) \forall i, j, k \in R
\end{align*}
(4)

Approaching this problem directly by means of optimization is considerably complex: while route enumeration can be formulated exactly through very simple mixed integer linear programs, the independence constraints are highly non-linear in nature, thus requiring specific solvers and reducing any guarantees of finding a definitive optimal solution.

We solve this problem through a graph theoretical reformulation, which manages to exactly capture the nature of the independence constraints of (4) through an hypergraph and its specific composition rules. Thanks to this approach, we are capable of computing route sets according to the conditions i)-iii).

While formal detailing of the methodology and its inner workings is left to an upcoming publication, in this work we focus on the impact of enumerating route sets following the aforementioned independency criteria, and how this relates to our hypothesis, i.e. how maximizing the NR and RI sets yields higher information content in observability solutions. To form this comparison, we measure the information content embedded in route sets generated through different mechanisms. Specifically, we employ three different route set generation techniques to construct the link-to-route incidence matrix \( A = [a_{ij}] \) needed as an input for link inference full observability problems, and then proceed to solve the full observability problem through Castillo’s pivoting procedure, obtaining thus the observability matrices \( \Omega \in \mathbb{R}^{[|k|] \times [|R|]} \). The three route set generation techniques we employ are the standard K-Shortest Path algorithm (KSP), our previously developed K-Independent Shortest Path (KISP) approach and the newly introduced hypergraph (HG) approach.

We measure the quantity of information embedded in each full observability solution through the following metric, measuring the information ranking of full observability solutions based on the observability matrices \( \Omega \):

\[ ||\Omega||_{F} \cdot rk(\Omega) \]

where \( rk(\Omega) \) represents the matrix rank operation.

This metric assigns a positive scalar value to any observability matrix \( \Omega \) and, as we experimentally showed in our previous works (Viti et al., 2014, Rinaldi et al., 2015), the higher the metric value the higher the information content embedded in a given full observability solution.

Based upon these considerations, we compare the three route set generation approaches across four simple networks, following three consecutive steps:

1) Three route sets \( R_{\text{KSP}}, R_{\text{KISP}} \) and \( R_{\text{HG}} \) are generated.

2) For each route set, corresponding full observability solutions \( \Omega_{\text{KSP}}, \Omega_{\text{KISP}} \) and \( \Omega_{\text{HG}} \) are computed following Castillo’s Pivoting procedure.

3) Each full observability solution is assigned a numerical value following (5)

It is to be noted that for both KSP and KISP Castillo’s Pivoting procedure yields non-unique results, depending on the order with which the routes composing the respective sets are included in either independent or dependent set. As for our previous approaches, we capture this stochasticity by repeating the procedure for 300 different ordering permutations. The newly introduced graph theoretical approach exhibits a systematically lower sensitivity to this stochasticity effect, and thus a single result is, as will be shown later, sufficient.

**Experimental results**

To verify whether route sets generated following our proposed independence constraints are indeed yielding high information in terms of full observability solutions, in this Section we collect and discuss results obtained through four small networks, shown in Figure 2. These networks, already studied (among others) in recent works of Castillo, provide a sufficient amount of complexity in terms of route-link relationships, while still being algorithmically tractable.

Applying our new optimisation formulation, full observability solutions derived using the identified independent route sets are characterised by a systematically lower amount of independent variables with respect to the alternative approaches proposed in (Rinaldi et al., 2015) and in (Castillo et al., 2014). Respectively for the four networks of Figure 2 we found full observability solutions requiring 6, 10, 10 and 8. Rinaldi’s KISP algorithm, which is systematically the second best approach, yields instead 9, 11, 13 and 8. Hence, our new solution algorithm performs equally well only in the last case, while it outperforms KISP in all other cases.
Figure 2: toy networks used to study the impact of route sets in terms of partial observability

Figure 3: NSP metric results using the three route set generation methods (5)
In Figure 3 we showcase the results of our route set generation comparison for all four networks: rank distributions related to KSP (shown in red), KISP (green) and HG (black) are shown. Analysing these comparative results two main observations can be drawn:

1. The amount of information embedded in the HG-generated full observability solutions is consistently higher than that of the two enumerative strategies.
2. The density of information reached by the newly proposed approach is indeed such that the full observability solution can be computed through a one-shot procedure at little impact in terms of information content, reducing (if not, removing) the need for exploring multiple permutations in the pivoting procedure.

These observations are indeed in line with our initial hypothesis: determining the maximum independent route set yields very high information content from the perspective of partial observability, indeed maximizing the \([NR \cup RI]\) set.

**Outlook**

We formulated a general optimisation problem aiming at maximising the amount of non-redundant and redundant while informative information, and in the same time minimising the purely redundant one. Exact solutions rely on the construction of an opportune hypergraph. Algorithms have been shown to efficiently identify solutions for different toy networks. The identified routes sets have the desirable properties of finding through standard pivoting procedures very efficient solutions in terms of full observability, both in terms of number of located sensors, as well as consistently finding information-rich solutions.

**Bibliography**


Supply Chain Logistics & Methods
FB4: Freight Transportation
Friday 1:00 – 2:30 PM
Session Chair: Ronald Askin

1:00  Scheduled Service Network Design and Revenue Management with an Intermodal Barge Transportation Illustration
1Teodor Gabriel Crainic*, 2Ioana Bilegan, 3Yunfei Wang
1ESG, UQAM & CIRRELT, 2LAMIH, University of Valenciennes, 3University of Valenciennes

1:30  Load Commitment Policies for the Stochastic Advance Booking Problem for Truckload Trucking
Juliana Nascimento*, Hugo Simao, Warren Powell
Princeton University

2:00  Logistic Network Design for Daily Cyclic Truck Routes
1Ronald Askin*, +2Zhengyang Hu, 2Guiping Hu
1Arizona State University, 2Iowa State University
Scheduled Service Network Design and Revenue Management with an Intermodal Barge Transportation Illustration

Teodor Gabriel Crainic
CIRREL T and School of Management
Université du Québec à Montréal, Canada
Email: TeodorGabriel.Crainic@cirrelt.net
Yunfei Wang
Ioana C. Bilegan
LAMIH UMR CNRS 8201
Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes, France

1 The Tactical Planning and Revenue Management Issue

Intermodal freight transportation is generally defined as moving cargo, loaded into containers, by a series of at least two transportation modes, containers being transferred from one mode to the next at intermodal terminals, e.g., ports and rail yards (Bektaş and Crainic, 2008; Crainic and Kim, 2007; van Riessen et al., 2015; SteadieSeifi et al., 2014). Consolidation-based carriers, e.g., motor carriers, railroads and navigation companies, perform the largest share of intermodal transportation. Carriers aim to meet forecast shipper demand and requirements and maximize net profits, by setting up a resource and cost-efficient service network and schedule, the so-called tactical operations plan.

The scheduled service network design (SSND) problem class is the methodology of choice to build this tactical plan (Crainic and Kim, 2007). It selects the transportation services and schedule the carrier will operate and propose to shippers for the next season (e.g., six months). The schedule is built for a given schedule length (e.g., a week), which is then operated repeatedly for the duration of the season. SSND with resource management models also include the selection, allocation and routing of the main resources (e.g., vessels, locomotives, etc.) supporting the selected services (e.g., Andersen et al., 2009; Crainic et al., 2014).

Most service network design cases and models in the literature consider a single category of customers, making up what is generally identified as regular demand and is expected to represent most of what is serviced during any “normal” period. SSND mod-
els are thus set to minimize the cost of the system accounting for operating the selected services and resources plus, possibly, the cost of time for resources and freight. We take a different view and consider several categories of customers, tariffs and operation classes, and the maximization of the net revenue. Considering different types of customers and the possibility to capture more demand, or higher priced demand, by offering a different service network is rather new to the tactical freight transport planning literature. Moreover, the very few contributions focusing on revenue management and freight transportation (e.g., Bilegan et al., 2015; Wang et al., 2015) focus on the operational level, the tactical level being rarely envisaged (e.g., Crevier et al., 2012; van Riessen et al., 2015).

We aim to study the incorporation of revenue management considerations, usually tackled at the operational-planning level, into tactical planning models for intermodal consolidation-based freight transportation carriers. Our interest goes beyond the modeling and algorithmic challenges, to exploring the impact of this integration on the structure of the service network (e.g., should the carrier increase the offer of service - more departures or larger vessels - in order to later be able to capture spot demand?) and the selection of customer demands to service.

2 SSND-RRM Model Formulation

Let $G^{ph} = (N^{ph}, A^{ph})$ represent the physical network supporting the operations of the carrier, set $N^{ph}$ the intermodal terminals and set $A^{ph}$ the possible navigation movements between two adjacent ports. Each terminal $i \in N^{ph}$ is characterized by a vessel berthing capacity $Q_i$. The carrier operates vessels of various types, each type $l \in L$ being defined by its capacity $\text{cap}(l)$, travel time $\delta_{ij}(l)$ over arc $(i, j) \in A^{ph}$, and maximum number of available vessels $B_l$. The service network is to be set up for a schedule length discretized into $T$ periods, to answer the customer demand for transporting containers of type $\gamma \in \Gamma$ among the terminals in $N^{ph}$.

We model customer service and fare differentiation through a two-dimensional mechanism: business relationship and level of service requirements. The former principally addresses the contractual profile depending on the commitment to work with the carrier: regular customers with long-term contracts or understandings and demand $d \in D^R$ the carrier must satisfy, proportional-punctual customers present on the spot market for which one may decide on the proportion of their demand $d \in D^P$ to service, and full-punctual customers also present on the spot market but which, once selected, require that their complete demand $d \in D^F$ be transported. Two delivery types are defined with respect to the latter, slow and fast delivery reflecting the due times at destination requested by customers. One normally expects fares to reflect this differentiation, e.g., fast delivery requests would be priced higher than slow delivery ones. Each demand $d \in D^R \cup D^P \cup D^F$
is then characterized by the customer category \( \text{cat}(d) \), with associated fare class (delivery type) \( \text{class}(d) \) and fare amount \( f(d) \), and the number \( \text{vol}(d) \) of containers of type \( \gamma(d) \) available at the origin node \( \text{orig}(d) \in \mathcal{N}^\text{ph} \) at time period \( \text{in}(d) \), to be delivered to the destination node \( \text{dest}(d) \in \mathcal{N}^\text{ph} \) in period \( \text{out}(d) \) at the latest.

The formulation is defined on a space-time network \( \mathcal{G} = (\mathcal{N}, \mathcal{A}) \) capturing the time-dependency of the demand and the barge services. The node set \( \mathcal{N} \) is obtained by duplicating all physical nodes at all periods in the schedule length, while the set of arcs \( \mathcal{A} \) is the union of the set of holding arcs at terminals and the set of possible movements performed by barge services. A barge service \( s \in \mathcal{S} \) is defined in the space-time network by its physical origin and destination terminals, \( \text{orig}(s) \) and \( \text{dest}(s) \), ordered set of consecutive stops \( \eta(s) = \{\text{i}_k(s) \in \mathcal{N}^\text{ph}, k = 0 \ldots (K - 1)\} \) and associated barge-service legs \( \{\text{a}_k(s) = (\text{i}_k(s), \text{i}_{k+1}(s))\} \), leg travel time \( \delta_k(s) \), stopping length \( w_k(s) \) at terminal \( \text{i}_k(s) \), terminal arrival \( \alpha_k(s) \) and departure \( \gamma_k(s) \) times, and a total duration \( \delta(s) \).

Let \( \phi(s) \) be the fixed cost of setting up and operating \( s \in \mathcal{S} \), \( h(i, l) \) the idling cost of a vessel of type \( l \) at terminal \( i \), \( c_k(\gamma(d), l(s)) \) the carrying cost of a unit of demand of type \( \gamma \) by barge-service \( s \) on its leg \( k \), and \( \kappa(i, \gamma(d)) \) and \( c(i, \gamma(d)) \) the unit loading / unloading and holding at terminal \( i \) costs, respectively, for type \( \gamma(d) \) demand.

The decision variables are: \( y(s) = 1 \) if service \( s \) is selected, 0 otherwise; \( \xi(d) \in [0, 1] = \) fraction of demand \( d \in \mathcal{D}^\text{p} \) selected to be serviced; \( \zeta(d) \in \{0, 1\} = 1 \) if the demand \( d \) is selected to be serviced, 0, otherwise; \( z(l, i, t) = \) number of idle vessels of type \( l \) waiting at terminal \( i \) during time period \((t, t+1)\) to start on a new service; \( v(l) = \) total number of vessels used by the service plan; \( x(d, s, k) = \) volume of demand \( d \) transported by service \( s \) on its leg \( k \); \( x^{\text{in}}(d, s, k), x^{\text{out}}(d, s, k) = \) volumes of demand \( d \) loaded/unloaded on/from leg \( k \) of \( s \); \( x^{\text{hold}}(d, i, t) = \) volume of demand \( d \) on hold at terminal \( i \) during time period \((t, t+1)\).

The objective function of the SSND-RM model aiming to maximize the net profit is:

\[
\max \sum_{d \in \mathcal{D}^\text{p}} f(d)\text{vol}(d) + \sum_{d \in \mathcal{D}^\text{f}} f(d)\xi(d)\text{vol}(d) + \sum_{d \in \mathcal{D}^\text{f}} f(d)\zeta(d)\text{vol}(d) \\
- \sum_{l \in \mathcal{L}} \phi(l)(\text{B}_l - v(l)) - \sum_{s \in \mathcal{S}} \phi(s)y(s) - \sum_{l \in \mathcal{T}} \sum_{i \in \mathcal{N}^\text{ph}} h(i, l(s))z(l, i, t) \\
- \sum_{s \in \mathcal{S}} \sum_{k \in \eta(s)} \sum_{d \in \mathcal{D}} c_k(\gamma(d), l(s))x(d, s, k) - \sum_{l \in \mathcal{T}} \sum_{i \in \mathcal{N}^\text{ph}} \sum_{d \in \mathcal{D}} c(i, \gamma(d))x^{\text{hold}}(d, i, t) \\
- \sum_{s \in \mathcal{S}} \sum_{k \in \eta(s)} \sum_{d \in \mathcal{D}} \kappa(i, \gamma(d))(x^{\text{in}}(d, s, k) + x^{\text{out}}(d, s, k))
\]

The three first terms correspond to the potential revenue obtained from serving regular, proportional-punctual and full-punctual customers respectively, the following ones being the different costs that must be subtracted in order to compute the total net profit (i.e., cost/penalty for vessels belonging to the fleet but not used during the schedule length, fixed cost of setting up and operating barge services, holding cost for idle vessels at ports, container transportation, holding (at terminals) and loading/unloading costs).
The constraints (skipped for lack of space) of the model enforce the customer container-flows in the physical and service networks, the vessel capacities, the numbers of available vessels of all types, the terminal berthing capacities, and the resource conservation (design-balanced restrictions).

3 Solution Methods and Result Analyses

We generate instances reflecting the networks in northern France (and beyond in Benelux). The instances are of moderate dimensions to allow solving the corresponding models with CPLEX. Instances are characterized by different topologies for the physical network, fleet size, demand type distribution and volumes, and value ranges for service types.

The first set of experiments, performed using CPLEX, showed the 1) it is feasible to consider revenue management within service network design formulations for tactical carrier planning, and 2) this integration appears rewarding. Several insights were gained through this experiments, e.g., the evolution of cost (increase) and resource utilization (more small vessels, decreasing loading rates) with the increase in demand for superior service, and the significant impact of combining several customer types and fares not only for the net profit, but also on the efficient utilization of the fleet.

It also showed the limits of commercial MIP software for such formulations. We therefore developed a math-heuristic combining tabu search and adaptive large-neighborhood search ideas. The meta-heuristic builds on the problem characteristics in defining two sets of neighborhoods to modify, the sets of accepted full-punctual demands and the selected services, respectively. Within each set, neighborhoods are defined based on single or multiple add, drop and swap principles.

The core of the meta-heuristic then consists in a series of local improvements based on biased-random selection of neighborhoods. The performance of the neighborhood and the retained services and demands (in demand and service neighborhoods, respectively) are recorded and used in biased-random selection mechanism. Intensification and diversification phases complete the meta-heuristic.

Initial experimental results (this part of the work is still in progress) indicate that the current version of the meta-heuristic performs acceptably well. It displays an average of less than 10% gap with respect to CPLEX (for the same amount of CPU time) for the small-dimension instances CPLEX may solve to optimality. The meta-heuristic overtakes CPLEX as soon as dimensions increase, obtaining better solutions faster.

4 The Presentation

We will present the SSND-RRM model and the meta-heuristic, and discuss the insights provided by the experimentation campaign. Initial tests on instances that may be ad-
dressed with commercial MIP solvers yield interesting results, regarding the impact of service/fare differentiation on profits and system performance, and changes in the selected service network. The results on large problem instances will be covered in depth.

References


Load Commitment Policies for the Stochastic Advance Booking Problem for Truckload Trucking

Juliana Nascimento, Hugo Simao, Warren B. Powell

Department of Operations Research and Financial Engineering
Princeton University
email: powell@princeton.edu

December 26, 2016

Truckload motor carriers face the complex problem of managing fleets of drivers, observing hours of service commitments, managing driver productivity while also getting drivers home in time, while serving a highly stochastic set of loads. While the management of fleets of vehicles (and drivers) has received considerable attention for many years, largely overlooked is the problem of making commitments to loads offered in the future. It is not unusual to find long-haul truckload carriers where 50 percent of the loads are called in four or more days in the future, long before the carrier knows where drivers will be at that time. Our research addresses the problem of designing effective policies for making these commitments.

There has been an extensive literature on dynamic fleet management in the presence of uncertain demands. The earliest work in this setting was in the context of freight car distribution (Jordan & Turnquist (1983)). This work motivated a lengthy series of papers on the dynamic vehicle allocation problem (Cheung & Powell (1996), Powell & Godfrey (2002), Powell & Topaloglu (2005), Topaloglu & Powell (2006)) which was centered on dynamic programming formulations. This line of research focused on managing fleets of vehicles (see Crainic et al. (1993), Powell et al. (1995) and Yang et al. (2004) for extensive reviews). Powell (1987) and Powell et al. (1988) were some of the earliest papers to deal with the complexities of managing drivers. This work finally matured in Simao et al. (2009) which used approximate dynamic programming to optimize drivers, taking into consideration all the complex attributes of drivers (home domicile, equipment type, hours of service, single vs. team, border crossing capabilities).

None of these papers deal with the problem of optimizing the advance booking of loads, which is quite common in truckload trucking. All the prior models assume that we make decisions to move trailers, containers or drivers at time \( t \), and then learn information about new demands at time \( t+1 \). The models may not satisfy all the demands, but there is no explicit decision of whether to accept to move a load, which is difficult when the load is to be moved in the future.

In the truckload industry, a shipper (or broker) may call a carrier and offer a set of loads to be moved, identified by the day of pickup (typically the loads are all picked up on the same day in the future), along with their origin and destination. Typically, a “shipper” represents a manufacturer (or warehouse) so that all the loads have the same origin and different destinations, although this is not always the case. Major shippers, with loads moving from multiple origins to many destinations, will ask carriers to bid on the lanes they wish to serve. A shipper may offer a carrier loads in all the lanes in which the carrier has submitted a bid (which includes the freight rate), but the carrier is not required to accept all the loads every day. If the carrier does accept a load, then it also understands that there will be a service penalty if it is not moved on time.

We build on the driver management model presented in Simao et al. (2009), which handles driver characteristics at a high level of detail, but where loads evolve in strict time order. In this research, we begin by developing a much more detailed model of the load booking process, including capturing the full booking profile which captures how far each shipper books loads in the future. We then use this stochastic model as the basis of a stochastic lookahead model. Next, we use the ADP policy presented in Simao et al. (2009) to estimate the probability that a load will be accepted. We design different load acceptance policies that depend on this load acceptance policy. Finally, these policies are tested in a newly developed simulator which represents the stochastic base model that we are trying to optimize using our load acceptance policy.
1 The base model

The base model is run as a simulator that implements decisions about accepted loads, and the assignment of drivers to loads. These decisions are made using policies we describe later. Following Powell (2011)[Chapter 5], our model consists of state variables, decision variables, exogenous information variables, the transition function, and the objective function.

1.1 The state variables

The driver management model is described in detail in Simao et al. (2009), so we provide only the general framework. We let

\[ a = \text{The attributes of a driver, including location, estimated time of arrival, home location, equipment, fleet (e.g. owner operator, company driver or sleeper team), and hours of service, where } a \in A, \]

\[ R_{ta} = \text{The number of drivers with attribute } a \text{ at time } t, R_t = (R_{ta})_{a \in A}, \]

\[ b = \text{The attributes of a load, including pickup time, pickup location and destination, where } b \in B, \]

\[ D_{tt'b} = \text{The number of loads with attribute } b \text{ that have been offered and accepted by the carrier by time } t \text{ that need to be picked up at time } t', D_t = (D_{tb})_{t \geq t, b \in B}, \]

\[ D_t^x = \text{The number of loads remaining to be served after dispatch decisions have been made at time } t, \]

\[ \hat{D}_{tt'b} = \text{The number of loads with attribute } b \text{ that are offered to the carrier in the interval } (t-1,t) \text{ to be picked up at time } t'. \]

The state of our system at time \( t \) is given by \( S_t = (R_t, (D_t^x, \hat{D}_t)). \)

1.2 Decisions

We distinguish between decisions to dispatch drivers \( x_t \), and decisions to accept loads \( y_t \). The driver decisions are defined using

\[ d = \text{A decision that acts on a driver, which may be the decision } d = \phi \text{ to do nothing, or a decision } B \text{ to move a load with attribute } b \in B, \]

\[ D = \phi \cup B, \]

\[ D_t^x = \text{The set of possible decisions that can be used to act on a driver,} \]

\[ \delta_{tdb} = 1 \text{ if } d \in D_t \text{ corresponds to a load with attribute } b, 0 \text{ otherwise,} \]

\[ x_{tad} = \text{The number of drivers with attribute } a \text{ that we act on with a decision of type } d, \]

\[ x_t = (x_{tad})_{a \in A, d \in D}, \]

\[ y_{tt'b} = \text{Number of loads accepted at time } t \text{ to be moved at time } t' \text{ with attribute vector } b, \]

\[ y_t = (y_{tt'b})_{t \geq t, b \in B} \]

Decisions have to be made subject to constraints that each driver is assigned to do something (hold or move to a load), a load cannot be assigned to more than one driver, and we either have to accept or reject each offered load. These decisions are determined by policies which are functions that return a decision given the information in \( S_t \). Since we have driver assignment decisions and load acceptance decisions, we denote these policies by \( X^\pi(S_t) \) (for the driver assignment policy) and \( Y^\pi(S_t) \) (for the load acceptance policy).

1.3 The exogenous information process

For this work, the only exogenous information process we consider is the process of loads being called in. This is given by the random vector \( D_t \) that we introduced earlier (because it is part
of the state variable). We will have other sources of exogenous information such as travel time delays. We let \( W_t \) be the variable capturing all the information that first becomes known at time \( t \), including \( \hat{D}_t \).

The most important exogenous information process for this research is the stochastic load generation process \( \hat{D}_t = (\hat{D}_{t',v})_{v \geq t} \). We use historical data to fit a pre-book distribution, giving the probability that a load called in at time \( t \) needs to be picked up at time \( t' \). We fit the distribution for the number of loads tendered by each shipper on a day from an origin to all destinations, and then use spatial patterns to sample the destination for each load. The mean is adjusted using an adaptive learning model for the baseline that adjusts to day of week, week of year and an array of “special days” (holidays, days before holidays). An inverse transformation is used to ensure that we match the actual distribution. Special care is taken to ensure that distributions match historical distributions at different spatial levels (but centered on outbound volumes from a single region).

1.4 The transition function

The dynamics of drivers are captured by the attribute transition function (introduced in Simao et al. (2009)):

\[
M(a_t, d, W_{t+1}) = \text{Attribute transition function, where } a_{t+1} = a^M(a_t, d, W_{t+1}) \text{ is the attributes of a driver with attribute } a_t \text{ after being acted on by decision } d \text{ (described below), given exogenous information } W_{t+1}.
\]

The attribute transition function is a piece of logic that handles all the hour-of-service constraints, as well as random delays to drivers. This is described in detail in Simao et al. (2009).

In our algorithms, we make extensive use of the post-decision state variable, which is the state after decisions have been made, but before any new information has arrived. Let \( \overline{W}_{t:t+1} = \mathbb{E}\{W_{t+1}|S_t\} \) be our forecast of \( W_{t+1} \) given what we know at time \( t \). For the resource vector, we can then use \( \overline{W}_{t:t+1} \) instead of \( W_{t+1} \) to get the attribute of a driver after a decision has been made (for example, using forecasted travel times). We use this to find the post-decision resource vector

\[
R^*_x = (R^*_{ix})_{a \in A},
\]

where \( R^*_x \) is the number of drivers with expected attribute vector \( a \) after decisions \( x \) have been made at time \( t \). As new information arrives in \( W_{t+1} \), we adjust the attribute vector (such as recording that the truck may arrive late). Below we also need

\[
\delta_{ad}(a') = \begin{cases} 1 & \text{if decision } d \text{ acting on a resource with attribute } a \text{ produces a resource with attribute } a' = a^M(a_t, d, \overline{W}_{t:t+1}), \\ 0 & \text{Otherwise} \end{cases}
\]

Central to this research is the evolution of the loads. The load acceptance policy acts on the offered loads in \( \hat{D}_t \) (these are loads offered between \( t-1 \) and \( t \)). We let \( D^A_t \) be the loads accepted by the load acceptance policy out of the set of offered loads \( \hat{D}_t \), where \( D^A_t = (D^A_{t'v})_{v \geq t, b \in B} \) is the vector of loads with dispatch time \( t' \) and attribute vector \( b \). To simplify the notation, we write the process of the load acceptance policy acting on \( \hat{D}_t \) to produce the set of accepted loads as

\[
D^A_t = Y^x(S_t) \otimes \hat{D}_t.
\]

We next write the set of loads covered by the driver assignment process as \( D^C_t \) where an element of \( D^C_t \) given by \( D^C_{t'v} \) is the number of loads remaining to be served at time \( t' \) with attribute \( b \) after all assigned loads have been subtracted. Again to simplify notation we write this as

\[
D^C_t = X^x(S_t) \otimes D_t,
\]
We can now write the evolution of loads using

\[
D_t^c = \text{The loads remaining after the dispatch decisions have been made (this is the post-}
\]

\[
\text{decision load vector)},
\]

\[
D_t^c = D_t - D_t^c,
\]

\[
D_{t+1} = D_t^c + D_{t+1}.
\]

This section provides a very high level sketch of the transition function, which captures all the
dynamics of the problem. We model the complete system dynamics using the system model

\[
S_{t+1} = S^M(S_t, (x_t, y_t), W_{t+1})
\]

where \(x_t = X^\pi(S_t)\) and \(y_t = Y^\pi(S_t)\) (note that we have to determine \(y_t\) before finding \(x_t\)).

### 1.5 The objective function

Each decision \(d\) acting on a driver (usually to move a load) generates costs (e.g. moving empty)
and revenues (from moving the load). Of course, drivers can only move loads that have been
accepted by the load acceptance process. Let

\[
c_{tad} = \text{Contribution from acting on a driver with attribute } a \text{ with a decision } d, \text{ that may}
\]

\[
\text{involve moving a load. This contribution captures empty movement costs, loaded}
\]

\[
\text{revenues, and any service penalties from picking up or delivering loads early or late.}
\]

We denote the total contribution at time \(t\) resulting from a decision \(x_t = X^\pi(S_t)\) using

\[
C(S_t, x_t) = \sum_{a \in A} \sum_{d \in D} c_{tad} \cdot x_{tad}.
\]

Now let \(\pi^y\) represent the characteristics of the load acceptance policy \(Y^\pi(S_t)\), and let \(\pi^x\) be
the characteristics of the driver dispatch policy \(X^\pi(S_t)\). If we fix \(\pi^y\), the problem of finding the
best driver dispatch policy is given by

\[
F^{\pi^y} = \max_{\pi^x} \left\{ \sum_{t=0}^{T} \sum_{t=0}^{T} C(S_t, X^\pi(S_t)) | S_0 \right\}.
\]

given the transition function (1). Finally, the optimization problem that is the focus of this research
is to solve \(\max_{\pi^y} F^{\pi^y}\). Equation (2) combined with the transition function (1) represents what we
are going to call our base model. Our next challenge is to design our dispatch policy \(X^\pi(S_t)\) and
the load acceptance policy \(Y^\pi(S_t)\).

### 2 Designing policies

We are going to approach the design of policies by using lookahead models that start at a time
\(t\), and optimize over a horizon \((t, t + H)\). Ideally, the lookahead model is the same as the base
model for those time periods, but in practice, it is very common to introduce other modeling
approximations (e.g. differences in time steps, discretization of states, the model of the exogenous
information process). To avoid confusion between the base model and the lookahead model, we
use the same notation as the base model but introduces tilde’s on all the variables. Also, while we
are interested in optimizing the entire lookahead model, we are only going to use the decisions
recommended at time \(t\). Thus, our dispatch policy would be written

\[
X^\pi_t(S_t) = \arg \max_{x_t} \left( C(S_t, x_t) + \mathbb{E}_{\pi^x} \left\{ \max_{\pi^y} \mathbb{E}_{\pi^x} \left\{ \sum_{t'=t+1}^{t+H} C(S_{t'}, \tilde{X}_t^\pi(S_{t'})) | S_{t'}, x_t \right\} \right\} \right),
\]
where $\tilde{S}_{t, t' + 1} = S^M(\tilde{S}_{t'}, \tilde{x}_{t'}, \tilde{W}_{t, t' + 1})$. This problem is solved in Simao et al. (2009) using approximate dynamic programming, which produces a set of policies based on value function approximations

$$X^*_t(\tilde{S}_t) = \arg\max_{\tilde{x}_t} \left( C(\tilde{S}_t, \tilde{x}_t) + \tilde{V}_t'(\tilde{R}^*_t) \right),$$

$$= \arg\max_{\tilde{x}_t} \left( C(\tilde{S}_t, \tilde{x}_t) + \sum_{a' \in A} \sum_{a \in A} \sum_{d \in D} x_{tad} \delta_{tad}(a') \tilde{v}_{t' a'} \right)$$

where $\tilde{V}_t'(\tilde{R}^*_t)$ is value of being in the post-decision resource state $\tilde{R}^*_t$ at time $t'$ within the lookahead model that starts at time $t$. This value function is computed using approximate dynamic programming as described in Simao et al. (2009), which means we have to stop to do iterative learning of the value functions at each point in time $t$ during our simulation of the base model (2).

Once the value functions for the lookahead model have been calculated (this typically requires several dozen forward simulations), we then fix the value functions $\tilde{V}_t'(\tilde{R}^*_t)$ (which fixes the dispatch policy) and then perform a series of simulations (say, 20) from a sampled set $\tilde{\Omega}_t$ which gives sample realizations $\tilde{W}_{t'}(\tilde{\omega})$ for $\tilde{\omega} \in \tilde{\Omega}_t$. This gives us a set of dispatch decisions $\tilde{x}_{t'}(\tilde{\omega})$ for $t' \geq t$ and $\tilde{\omega} \in \tilde{\Omega}_t$. We also exploit the fact that these simulations of a fixed policy in the lookahead model can be done in parallel.

Our lookahead model uses known loads $\hat{D}_{t-1}$ (loads leftover from time $t - 1$) and $\hat{D}_t$ at time $t$ from the base model, which includes loads to be moved in the future (recall that $\hat{D}_t = (\hat{D}_{t' b})_{t' \geq t, b \in \mathcal{B}}$). For simplicity, assume that $\hat{D}_{t' b}$ is 0 or 1. The lookahead model combines the loads in $\hat{D}_{t-1}$ and $\hat{D}_t$ with simulated loads from the process $(\tilde{W}_{t'})_{t' \geq t}$. As we simulate our dispatch policy for the lookahead model generated at time $t$, we will simulate whether the offered loads in $\hat{D}_t$ (which are fixed at time $t$) are accepted by the dispatch policy or not. Let

$$\hat{y}_{t' b}(\tilde{\omega}) = \begin{cases} 
1 & \text{If the load with attribute } b \text{ to be moved at time } t' \text{ is moved under sample path } \tilde{\omega}, \\
0 & \text{Otherwise.}
\end{cases}$$

We note that the offered loads $\hat{D}_t$ are determined for a particular lookahead model at time $t$, so their presence can affect the value functions $\tilde{V}_t'(\tilde{R}^*_t)$. We can then create an estimate that an offered load will actually be moved using

$$p^{A}_{t' b} = \frac{1}{\hat{D}_{t' b}} \sum_{\tilde{\omega} \in \tilde{\Omega}_t} \hat{y}_{t' b}(\tilde{\omega}).$$

Now we can build a load acceptance policy $Y^\pi(S_t)$ using these estimated probabilities. We propose a simple policy function approximation (PFA), which is one of the four classes of policies (see Powell (2016), Section 8):

$$y_{t' b}^\pi(\theta) = \begin{cases} 
1 & \text{If } p^{A}_{t' b} \geq \theta^A, \\
0 & \text{Otherwise},
\end{cases}$$

where $\theta^A$ is our minimum acceptance probability. Then, once a load has been “accepted,” it is modeled in the dispatch problem with an additional bonus (on top of the normal reward) of $\theta^B$ to represent the penalty of finding an alternative supplier if the carrier is not able to move the load.

Our policy $Y^\pi(S_t; \theta)$ consists of the vector $y_{t' b}^\pi(\theta)$ over $t' \geq t$ and $b \in \mathcal{B}$. This is parameterized by the acceptance target $\theta^A$ and the committed load bonus $\theta^B$. We will summarize both offline and online algorithms for finding the best value of $\theta = (\theta^A, \theta^B)$. 
References


Logistic Network Design for Daily Cyclic Truck Routes

Extended Abstract

Zhengyang Hu\textsuperscript{1}, Ronald G. Askin (corresponding author)\textsuperscript{2} and Guiping Hu\textsuperscript{1}
\textsuperscript{1}Department of Industrial and Manufacturing Engineering, Iowa State University, Ames, IA
\textsuperscript{2}School of Computing, Informatics, and Decision Systems Engineering, Arizona State University,
Tempe, AZ 85287, ron.askin@asu.edu

Key Words Physical Internet; network design; truckload transportation

1. Introduction

Hub or relay based logistics networks for long haul trucking have been proposed for improving truck driver work-life balance and other efficiencies. Recent discussion of the Physical Internet has renewed interest in such systems. In this paper a mixed integer programming decision model is proposed for hub network design and routing to support driver residency. The model determines the locations and capacities of hubs for the network to meet a set of source to destination shipping demands. The model minimizes fixed operating costs, loaded and deadheading travel cost and penalties for driver routes that exceed a daily tour. A case study based on the highway grid for the Western United States is used to demonstrate the mathematical model and operational practice.

According to the U.S. Department of Transportation, over $12B of freight is shipped by truck annually in the United States, accounting for over 60\% of total shipped weight and value. These truck shipments account for over 40\% of the ton-miles of shipped freight. Two thirds of these ton-miles are for shipments over 250 miles and over 10\% are for shipments in excess of 2,000 miles. Clearly system efficiency and sustainability are of critical importance. Efficiency refers to system costs and profitability. Sustainability in this context refers to both operational and environmental sustainability.

Trucking can be classified into two types: (i) full truckload trucking (TL), generally with loads in excess of 10,000 lbs., and (ii) less than truckload trucking (LTL). For TL, Point-to-Point (PtP) shipping occurs from large suppliers to major customers when demand volume and desired frequency warrant direct TL shipments. However, in many instances smaller loads are consolidated into full TL loads at local hubs to facilitate cost efficiency. Shipments however are rarely temporally symmetric and TL problems have generally focused on how to minimize the unloaded deadhead movement. Routes are often created with multiple legs, even beyond long PtP shipments, causing drivers to be away from home for extended periods. This method has substantial benefit for companies and customers in terms of cost and lead time, but it is unaffordable both mentally and physically for truck drivers. Long tour lengths increase driver turnover. Moreover, Meller and Ellis (2012) indicated that trucks are completely empty 25\% of total time, and utilization is only 57\% for the other 75\% of loaded time. Mele (1989) observed
that turnover rate ranges from 85% to 110% per year in the TL industry and that problem persists today creating a shortage of long haul truckers. By contrast, the turnover rate for local drivers with daily tours is significantly lower. Safety may also be an issue as drivers who are assigned away from home for a long time incur diet and sleep problems that can increase the accident rate.

Hub, or Relay Networks (RNs) have been considered as an alternative to the traditional long-haul TL system. In a RN as considered here, a system of hubs is established along an interstate traffic network such that the distance between hubs is less than a half-day drive. Drivers are stationed at hubs (presumably near their residence) and make a round trip (or circuit), ideally starting and ending up at their home station each day. Loads (trailers) are exchanged at the hub and continue on their route for longer trips. This operational procedure potentially allows a load to be in virtually continuous transit rather than having to sit idle during driver break periods. It does however have the ability to add distance to the trip relative to PtP shipments. Decisions are thus where to place the hubs, the capacity of each hub, size of the driver workforce and the routing of loads. Taylor et al. (1999) considered the hub problem from three aspects: (i) customer viewpoint, (ii) company viewpoint, (iii) driver viewpoint. They concluded that for long trips an effective dispatching scenario can improve driver work life, meet shipper’s efficiency goals and allow merchandise to be delivered on time without damage.

Recent discussion of the Physical Internet (PI) (Montreuil (2011)) has increased interest in relay systems. Meller et al. (2012) compared the conventional dispatching model where a single driver picks up loads at their source point and delivers them to their destination with a PI hub model.

The RN design problem is a variant of the multi-facility location problem (Love [1974]). Efroymson and Ray (1966) presented discrete plant location as a mixed integer programming (MIP) problem, determining the number, locations and sizes of plants. The difference is that more factors such as deadhead cost and hub exchanges are considered in this paper. Ebery et al. (2000) presented a MIP formulation based on capacitated p-hub median problem.

Üster and Maheshwari (2007) considered strategic network design for multi-zone truckload shipments. They formulated a model in order to choose the location of each Relay Node/Point (RP) and assignment of non-RP nodes to an RP node to include the original origin and destination. Üster and Kewcharoenwong (2011) modified the model and applied Bender’s Decomposition to solve it. The relay network addressed in those papers is very similar to the problem addressed in this paper however the model and problem definition differ in several dimensions. As is done in packet-routing networks, this paper does not restrict source-destination paths to be unique, but does impose capacity limits on hubs and directly models capacity costs. Deadheading is estimated by arc and its costs are included in the objective. Extended tours that prevent the driver returning home each day are modeled and allowed subject to a penalty cost where effective instead of utilizing hard constraints. As a result the number of binary variables is significantly reduced and a more complete economic model of the interhub network and operations are developed. This paper does however restrict focus to interhub traffic, assuming all regional transit is handled separately and therefore effectively originates or terminates at a major hub location. In practice this would mean a major port or urban area. In
section 2, the specific problem formulation is provided. Computational results, algorithms and extensions are described in section 3.

2. Problem Definition and Formulation

Our objective is to define the minimum cost set of capacitated flow-through locations on a network $G=(N,A)$ of node set $N$ and arcs $A$, and the accompanying routes, to serve a known set of source destination pairs $(i,l)$ with associated flow volume (weight) $w_{il}$. Each arc $(j,k) \in A$ represents a leg of a trip for a driver and has associated travel time $d_{jk}$. Loaded trucks cost $t_{jk}$ per unit travel time and deadhead costs are $z_{jk}$ per time. We assume the set of nodes $n \in N$ that can serve as possible hubs is known along with the allowable capacity and associated fixed annual depreciation and operating cost of each option. Trips with time in excess of $\alpha$ incur a penalty for not allowing the driver to return to his/her home base. Arcs are assumed to be bidirectional but not necessarily of equal transit time. For the computational tests we employ a rectangular grid and also the Western United States highway system with key cities as potential hub locations. In accord with Road Safety Authority (RSA), the longest driving time without resting has to be less than 5 hours. The fixed cost and capacity of option $z$ hub at site $j$ are denoted $f_{jz}$ and $c_{jz}$ respectively. Parameters $p_{jk}$ and $m_{jk}$ capture the overtime driving distance and cost per travelling time per load. These are only nonzero for links exceeding the allowed half-shift one-way travel time. Decision variables are $X_{ijkl}$, $Y_{jz}$ and $K_{jk}$. The $X_{ijkl}$ represent the proportion of demand from $i$ to $l$ using $j$ and $k$ as adjacent stops. $Y_{jz}$ is a binary indicator of whether a size $z$ relay hub is built at node $j$. Minimum deadhead travel is given by $K_{jk}$. As cities could support multiple sites we do not restrict the model to a single hub per node, but that can easily be accommodated in the model. The general model formulation then becomes:

Minimize $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} w_{il} \cdot X_{ijkl} \cdot (d_{jk} \cdot t_{jk} + p_{jk} \cdot m_{jk}) + \sum_{j=1}^{n} f_{jz} \cdot Y_{jz} + \sum_{j=1}^{n} \sum_{k=1}^{n} z_{jk} \cdot d_{jk} \cdot K_{jk}$

st:

$$\sum_{k=1}^{n} X_{ikl} - \sum_{k=1}^{n} X_{kil} = 1 \quad \forall {i,l} \quad (1)$$

$$\sum_{j=1}^{n} X_{ijl} - \sum_{j=1}^{n} X_{ijl} = 1 \quad \forall {i,l} \quad (2)$$
The objective function minimizes total cost of loaded transportation, overtime penalty, fixed hub operation, and deadhead travel. Constraints (1) through (3) enforce network flow constraints beginning at sources, ensuring conservation into and out of transit hubs and ending at destinations for each flow. Constraints (4) and (5) ensure hubs are constructed with sufficient capacity. Constraints (6) and (7) determine minimum deadheading based on the routing imbalance. The final constraints define variable types and ranges.

3. Computational results

Thirty-seven potential urban hub locations along the highway system in the Mountain and Western time zones of the US were considered in the study. Asymmetric travel times were obtained from [14] and annual truckload demands from [15]. For computational testing deadhead cost and penalty cost were assumed to be proportional to hourly driving cost and multiple hub options per node were considered.

A small model with 6 flows along the West Coast was first tested. Computations were performed on an Intel I7 processor. One small, four medium and two large hubs were built yielding deadhead cost of $2.9 \times 10^7$, transportation cost of $3.1 \times 10^8$ and fixed cost of $4.1 \times 10^8$. Extra travel due to hubs relative to point-to-point deliveries was $1.2 \times 10^7$.

A larger experiment tested up to 90 flows between the largest cities over a range of relative cost ratios. Various restrictions on the maximum length (days) of driver tours were also tested. All problems solved to optimality rather quickly. Depending on maximum allowable trip distance, solutions varied from less than a minute to 30 minutes with a standard solver. The ratio of
transportation cost to fixed facility cost and the ratio of hourly overtime penalty to regular transportation cost impacted the number and location of hubs. Additional testing is on-going to determine the range of parameter values that solve quickly.

Future extensions to the model include a two-stage stochastic optimization approach that constructs hubs based on the distribution of flow scenarios and then dynamically routes daily orders. The performance of the L-shaped Bender’s Decomposition Algorithm for solving large two-stage instances will be examined.

References

SUPPLY CHAIN LOGISTICS & METHODS
FC4: MOVEMENT AND MATERIALS CONTROL IN SUPPLY CHAINS
Friday 2:45 – 4:15 PM
Session Chair: Rajan Batta

2:45 Dynamic Capacity Logistics and Inventory Control
Satya Malladi*, Alan Erera, Chelsea White III
Georgia Institute of Technology

3:15 Planning the Fuel Supply to Gas Stations According to the Concept of Carrier-Managed Inventory - an Optimization Approach
Paweł Hanczar*
Wroclaw University of Economics

3:45 Exploration of Strategies to Form Convoys to Facilitate Effective Movement of Items
Rajan Batta*, Azar Sadehgheijad Barkousaraie, Moises Sudit
University at Buffalo, The State University of New York
Dynamic Capacity Logistics and Inventory Control

Satya S. Malladi*, Alan L. Erera*, and Chelsea C. White III*

*School of Industrial and Systems Engineering, Georgia Institute of Technology

1 Introduction

We investigate the logistics of distributed production-inventory systems that allow inventory and/or portable production capacity to be relocated. The mobile modular production problem (MMPP) we study in the current paper seeks the optimal decisions of module movement, and replenishment in every time period when faced with uncertain location-wise demands for a product.

The emergence of reconfigurable, mobile, decentralized/distributed manufacturing has generated an interest in the manufacturing industry in the recent years ([22, 23, 10, 24, 38, 37, 19, 20, 6, 32]), particularly, the success of Bayer [11], Pfizer [28], and Novartis [26] in creating prototypes of portable pharmaceutical production units. Amazon recently filed a patent for mobile 3D printing delivery trucks [1]. Such systems are characterized by transformability: “scalability, adaptability (modularity, universality, compatibility), and mobility” ([22, 38]).

![Examples of mobile modular technology](image)

Figure 1: Examples of mobile modular technology

Literature on capacitated single location inventory control ranges from [12] and [13] through [35, 21] (computing optimal policies), to [9, 4, 27, 17, 3] (exploring variable (uncertain) capacity). In the regime of joint capacity and inventory decision-making at a single location, Angelus and Porteus [2], [7, 31], and [30] are relevant contributions.

[38] deals with the MMPP from the dynamic facility location problem (DFLP) perspective with deterministic demands. [18, 14, 25] present related versions of the DFLP. Mobile facility routing [16] handles mobile capacity across multiple locations. However, the problem of jointly managing shared mobile capacity and inventory over all the locations has not addressed so far.

We model the MMPP as an MDP. We assume that a set of locations is given (see Clausen et al. [10] for approach). We seek to a) quantify the value addition by reconfigurable system over
a conventional system with fixed capacity plants and b) solve the MMPP close to optimality. A value addition study using real options analysis can be found in [38].

Due to state and action space explosion, the MMPP faces computational intractability. Hence, we use approximate dynamic programming techniques to obtain good heuristics for the problem at hand. We develop heuristics based on the bounds we propose. See [29, Ch. 10] for details. We also work with rollout heuristics using the fixed system as a guidance mechanism. For a rollout heuristic to be effective, the computation of the guidance mechanism, the fixed system, must be tractable [34, 5, 36, 33], which is another challenge we deal with. Goodson et al. [15] present a very useful framework of rollout policies.

2 Methodology

We consider a distributed production-inventory system with $L$ locations and $M$ portable production modules. At each decision epoch, we assume the decision maker (DM) knows the current demand forecast, inventory level, and production capacity at each location. This production capacity is made up of fixed capacity and portable capacity. The problem objective (of the DM) is to determine an optimal inventory and module relocation and replenishment policy in order to minimize the finite horizon expected total cost criterion. Demands at the locations are realized. Based on these realizations and perhaps other data, the demand forecast is updated prior to the next decision epoch. The chronology of events within period $(t, t+1)$ is as follows:

- Initial State
- Send modules
- Receive modules
- Produce
- serve demand
- incur cost
- New State

2.1 Problem Formulation

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>State (${u_i(t)}, {s_i(t)}$)</td>
<td>$u_i(t) =$ # modules at $i$ at the beginning of period $t$</td>
</tr>
<tr>
<td></td>
<td>$s_i(t) =$ inventory position at $i$ at the beginning of period $t$</td>
</tr>
<tr>
<td>Action (${x_i(t)}, {w_{ij}(t)}$)</td>
<td>$x_i(t) =$ # units produced at station $i$ in period $t$</td>
</tr>
<tr>
<td></td>
<td>$w_{ij}(t) =$ # modules moved from station $i$ to $j$ in period $t$</td>
</tr>
<tr>
<td>Uncertainty ($D_i$)</td>
<td>$D_i(t) =$ demand random variable at station $i$ in period $t$</td>
</tr>
<tr>
<td>Parameters</td>
<td>$p_i =$ backordering cost per unit per period at site $i$</td>
</tr>
<tr>
<td></td>
<td>$h_i =$ holding cost per unit per period at $i$</td>
</tr>
<tr>
<td></td>
<td>$c_i =$ production cost per unit at $i$(set to zero, $c_i = 0$)</td>
</tr>
<tr>
<td></td>
<td>$M =$ module fleet size</td>
</tr>
<tr>
<td></td>
<td>$I =$ inventory storage capacity at each station</td>
</tr>
<tr>
<td></td>
<td>$C' =$ capacity of each module, $C_i =$ fixed capacity at $i$</td>
</tr>
<tr>
<td></td>
<td>$T =$ length of the horizon, $T = {1, 2, \ldots, T}$</td>
</tr>
<tr>
<td></td>
<td>$L =$ # locations, $L = {1, \ldots, L}$</td>
</tr>
<tr>
<td></td>
<td>$K_{ij} =$ movement cost per module from $i$ to $j$</td>
</tr>
<tr>
<td></td>
<td>$l_M =$ module movement lag</td>
</tr>
</tbody>
</table>

Table 1: Notation

We make the following assumptions on the framework. At each decision epoch $t$ we assume the decision-maker (DM) knows ($u(t), s(t)$). Replenishment occurs instantaneously. Demand is described by known location-specific independent distributions. The location-wise 1 period inventory cost is non-negative, convex and divergent in order. The state spaces of $D_i(t)$, $\forall i$, are finite and the inventory state space is countable.
The feasible actions are determined by the following constraints:

\[
\sum_{i=1}^{T} u_i(t) + \sum_{\tau=1}^{l_M} I_{t>\tau} \sum_{i,j} w_{ij}(t-\tau) = M,
\]

\[
0 \leq x_i(t) \leq Gu_i(t+1), \quad \forall i \in \mathcal{I}, \forall t \in T \setminus \{T\}.
\]

The state dynamics are given by:

\[
s_i(t+1) = s_i(t) + x_i(t) - D_i(t) \quad \forall i \in \mathcal{I}, \forall t \in T \setminus \{T\}
\]

\[
u_i(t+1) = u_i(t) - \sum_{j \in \mathcal{I}} w_{ij}(t) + I_{t>l_M} \sum_{k \in \mathcal{I}} w_{ki}(t-l_M) \quad \forall i \in \mathcal{I}, \forall t \in T \setminus \{T\}
\]

i.e., \( \xi(t+1) = f(\xi(t), a(t, \xi(t)), D(t)) \forall t \in T \setminus \{T\} \)

Immediate cost accrued during period \( t \) i.e., \((t, t+1)\) is:

\[
C_t(\xi(t), a(t, \xi(t)), D(t)) = \sum_{i,j \in \mathcal{I}} K_{ij} w_{ij}(t) + \sum_{i \in \mathcal{I}} \left\{ c_i x_i(t) + h_i(s_i(t) + x_i(t) - D_i(t))^+ \right. \\
+ p_i(D_i(t) - s_i(t) - x_i(t))^+ \} \forall t \in T.
\]

Optimality Equations: Let \( V_t(\xi(t)) \) be the optimal cost-to-go function at time \( t \).

\[
V_t(\xi(t)) = \min_{a(t, \xi(t)) \in \mathcal{A}(\xi(t))} \mathbb{E}_D \left[ G_t(\xi(t), a(t, \xi(t)), D(t)) \right. \\
+ \beta V_{t+1}(f(\xi(t), a(t, \xi(t)), D(t))) \] \forall \xi(t), \forall t \in T \setminus \{T\}
\]

\[
V_T(\xi(T)) = 0 \forall \xi(T)
\]

where \( D(t) = \{D_i(t)\} \).

The above formulation gives a clear representation of the problem, capturing the characteristic sequential structure and uncertainty. However, it is plagued by state and action space explosions which make it intractable to solve for the exact solution (using either backward induction or LP method) of real-sized instances.

### 2.2 Bounds

The upper bounds on the optimal cost are:

1. **The fixed system (UB):** The conventional system with fixed capacity plants of equivalent total capacity allocated optimally among locations (mobile modularity is lost).

2. **Value function approximations:** Approximate value iteration initialized with the fixed system’s optimal cost function and updated based on simulated trajectories of states; the blending step to derive the current estimate of the value function can be:
   - constant, leading to VFA(const)
   - 1/ iteration number, leading to VFA(1/n)
   - 1/ number of visits to the state until now, leading to VFA(1/nVisits).

3. **Myopic Rollout Policy:** This policy implements the optimal policy for the one period problem at every decision epoch. Future cost is approximated by 0.

The lower bounds are:

1. **The most flexible system (LB):** The system in which the optimal configuration of total initial inventory and modules is attained instantaneously at no cost before replenishing.
2. The unlimited capacity system (UC): The system in which there is unlimited capacity at every location at every epoch gives a lower bound. The most flexible system is a tighter lower bound.

3. Perfect Information Relaxation (PIR) (for finite horizon): Non-anticipativity constraint is relaxed as demand realizations are known before decision-making.

2.3 Rollout Heuristic

We propose rollout strategies by relying on the bounds derived above. Using the intuition that the optimal cost lies between upper and lower bound, we approximate the future cost in successive approximations by a convex combination of the cost of the fixed system and the cost of holding unlimited capacity. This can be translated into the following strategy:

"Make inventory replenishment and module relocation decisions assuming that, from the next period onwards,

- w.p. $\lambda$, the best capacity configuration can be attained without cost and lag over the rest of the horizon and
- w.p. $1 - \lambda$, mobile modularity is lost (system is fixed) for the rest of the horizon."

$\lambda$ is a tuning parameter here.

This problem decomposition produces a dramatic increase in model tractability. However, as the state space is exponential in the number of locations, $L$, we need to explore techniques that do not rely on maintaining the value function over the entire state space.

2.4 ILP Approach

We model the single period MMPP as an integer linear program (ILP). For any initial state $\{(q_i), \{s_i\}\}$ of the problem, we introduce auxiliary variables, $z^k_i$, to handle the max terms in the objective function. The corresponding ILP is given by:

$$\text{ILP(MMPP1)} \min \sum_{i=1}^{n_s} \left[ \sum_{j=1}^{n_s} K_{ij} w_{ij} + \sum_{k=1}^{M} p_k h_i(y_i - z^k_i) + b_i(z^k_i - y_i) \right]$$

Subject to:

- $y_i - z^k_i \geq 0$, $\forall k \in \{1, \ldots, M\}, i \in \{1, \ldots, n_s\}$
- $z^k_i \geq -d^k_i$, $\forall k \in \{1, \ldots, M\}, i \in \{1, \ldots, n_s\}$
- $z^k_i \geq 0$, $\forall k \in \{1, \ldots, M\}, i \in \{1, \ldots, n_s\}$
- $y_i \geq s_i \forall i \in \{1, \ldots, n_s\}$
- $C' \left( -\sum_j w_{ij} + \sum_i w_{li} \right) - y_i \geq -s_i - C' q_i \forall i \in \{1, \ldots, n_s\}$
- $-\sum_j w_{ij} \geq -q_i \forall i \in \{1, \ldots, n_s\}$
- $y_i, w_{ij} \in \mathbb{Z}^+, \forall i, j$
- $z^k_i \in \mathbb{Z}^+ \forall k, \forall i$.

Here $y_i$’s are the order-up-to levels of inventory.

Chen et al. [8] prove that the newsvendor problem (one period inventory control problem) is integral under linear ordering, backordering and holding costs. We prove that ILP(MMPP1) is integral for $C' = 1$, and $T = 1$, with integral outcomes of demand $d_k$, $\forall k$, and integral $s_i, q_i$, $\forall i$.

In addition, we also prove that when $K_{ij} = K$, $\forall i, j$, the integrality constraints on $\{z^k_i\}$, and $\{y_i\}$ in ILP(MMPP1) are redundant (the resultant $y_i$ and $z^k_i$ are integers even if $w_{ij}$s are allowed to be non-integral).
3 Computational Analysis

We have performed computational experiments for the following two specific versions of the MMPP:

1. \( L = 2, C_i = 0 \) \( \forall i, l_M = 0, T > 1 \). We study the effect of varying the capacity per module, \( C' \), and the total fleet size, \( M \), and the length of the horizon, \( T \) on the performance of the bounds and the heuristic. We work with the simplest heuristic for these experiments with \( \lambda = 0 \), essentially, make decisions assuming that there no mobile modularity from the next period onwards.

2. \( L > 2, C_i = 0 \) \( \forall i, l_M = 0, T = 1 \). We study the effect of varying the capacity per module, \( C' \), and the total fleet size, \( M \), the number of locations, \( L \) on the efficiency of this approach.

3.1 \( L = 2, T > 1 \)

We have generated a set of 1280 instances by varying \( C' \), \( T \), \( M \) and the demand distributions. We compare the bounds and the heuristics w.r.t. OPT for this set.

\( \text{UB}_\text{min} \), the cost of the best fixed system is 5% higher than OPT. \( \text{LB} \) (0.93 times OPT on average) and \( \text{UB}_\text{min} \) improve with an increase in \( Y \), (see Figure 2a). \( \text{PIR} \) (0.218 times OPT on average) produces lower bounds of poor quality that deteriorate with an increase in system size, \( Y \). \( \text{UB}_\text{max} \) which averages at 8.4 times the OPT. \( \text{UB}_\text{max} \) increases rapidly with system size, \( Y \)

![Figure 2: Performance of bounds](image)

(a) \( \text{UB}_\text{min} \) vs. LB

(b) \( \text{UB}_\text{min} \) vs. \( \text{UB}_\text{max} \)

(see Figure 2b. The myopic rollout policy (RM) performs very poorly compared to the optimal policy, costing about 2.4 times the optimal policy on average, suggesting the importance of good estimation of the future cost.

<table>
<thead>
<tr>
<th>VFA (1/nVisits)/OPT</th>
<th>VFA (constant)/OPT</th>
<th>VFA (1/n)/OPT</th>
<th>RF/OPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.031</td>
<td>1.013</td>
<td>1.044</td>
<td>1.008</td>
</tr>
</tbody>
</table>

Table 2: Heuristics w.r.t OPT on average

In Table 2 VFA with constant blending coefficient (\( \alpha = 0.5 \)) and the fixed rollout policy outperform the other heuristics from Figure 3b. Let \( \eta \) be the average number of modules moved per period, averaged over those instances which induce movement under the policy being implemented. The optimal policy induces movement of modules in 572 out of 1280 instances, and the rollout policy induces movement of modules in 599 out of 1280 instances. Figure 3a shows that the fixed rollout policy shows close behavior with the optimal policy on average.

In general, the computation of the UB and hence the rollout strategy (takes about 1/20th of time taken by the OPT). VFA takes about 1/4.5th the time taken by the OPT on an Intel Xeon E5345 (Quad core, 2.33 GHz, 8MB), with 12 GB RAM.
<table>
<thead>
<tr>
<th>$T$</th>
<th>$\eta_{OPT}$</th>
<th>$\eta_{RF}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.2094</td>
<td>0.2038</td>
</tr>
<tr>
<td>7</td>
<td>0.1469</td>
<td>0.1447</td>
</tr>
<tr>
<td>10</td>
<td>0.1037</td>
<td>0.1010</td>
</tr>
<tr>
<td>15</td>
<td>0.0689</td>
<td>0.0670</td>
</tr>
<tr>
<td>Overall</td>
<td>0.1322</td>
<td>0.1291</td>
</tr>
</tbody>
</table>

(a) Performance of Fixed Rollout

![Figure 3: Performance of heuristics](image)

We have worked with a second set of instances with 720 instances each for $C' = 1, 2, 3$. The performance of bounds and heuristics w.r.t. the lower bound have been compared. While the value function approximations still perform well, the rollout strategy with $\lambda = 0$ produces policies in which the number of module movements reduce drastically with $C'$. This behavior can be explained as follows: unlike the optimal strategy, this strategy does not account for the future benefit of holding higher capacity at the new location. The immediate benefit is not a good enough incentive to move the modules often, especially by losing higher ($C'$) capacity at some location, leading to fewer module movements when $C'$ is greater than 1 but demand still occurs in units of 1. Hence, it is reasonable to work with $C' = 1$ in most situations.

3.2 $L > 2, T = 1$

The CPU times for computational experiments to solve ILP(MMPP1) optimally are presented.

<table>
<thead>
<tr>
<th># Locations, $L$</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU time (s)</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>34.61</td>
</tr>
</tbody>
</table>

Table 3: Efficiency w.r.t. $L$ for $C' = 3$

As capacity per module is changed proportionally with demand, the computation times do increase, but by less than 10% always.

4 Conclusions and Future Work

In this paper, we have quantified the value addition of mobile modular production systems. We have proposed heuristics and bounds as the exact computation of the optimal cost-to-go function and the optimal policy are expensive.

Future work may focus on improved heuristics that scale well with the number of locations and the number of modules. Also, spatial non-uniformity and temporal nonstationarity enable extracting highest value from mobile modular systems. Better models of both features would enhance the quality of the solution techniques provided here. A fleet management perspective of the multilocation, multimodule problem allows the application of networks, which would tackle the problem at larger instance sizes. We are working on this approach with the ILP technique. Last but not the least, an analysis of the mobile modular problem juxtaposed with the inventory transshipment problem would establish when which system would be effective. We are currently working on allowing inventory transshipment also in the MMPP framework with nonstationary demands in order to evaluate when module relocation is preferred over inventory transshipment.
5 Bibliography


Planning fuel supply to gas stations according to the concept of carrier-managed inventory - an optimization approach

Introduction

The vendor managed inventory (VMI) concept has many different variants in business practice. In the case of fuel distribution to a gas station network, VMI co-operation is often referred to as carrier managed inventories (CMI), as the transportation company takes over the responsibility of shipment planning and delivery dates.

Distribution companies expect the carrier not only to deliver specialized transport services, but also to develop end-to-end delivery plans to achieve key performance indicators (KPIs). On the one hand, companies that own a network of gas stations expect transport companies to operate in accordance with CMI as a prerequisite for cooperation. On the other hand, transport companies accept this approach, realizing that taking over additional responsibilities will result in higher utilization of the transport fleet and will, in the future, make it difficult for the distribution company to change the provider of transport services.

Transport companies deciding to provide services in accordance with the CMI concept face significant challenges in the preparation of plans to maintain the appropriate levels of indicators set by the client. Depending on the scope of these requirements, route planning can be implemented in an intuitive way, based on expert knowledge and the planner’s individual experience. Unfortunately, full use of the potential offered by CMI is only possible after applying advanced planning methods such as discrete optimization. These methods are also necessary when the number of KPIs and the complexity of their calculation increase.

In addition, the CMI concept gives transport companies the ability to limit the number of vehicles needed to deliver supplies to a given gas station network. High variability in the number of vehicles needed to deliver supplies in subsequent periods is very detrimental to the use of fleet capacity.

In this paper, the problem of determining the routes that occur in the process of planning fuel supplies to a gas station network was discussed. The most commonly encountered KPIs which are used in the cooperation between a carrier and their contractor were studied. The main part of the paper contains the proposal of a decision model for supply planning through a 7-day time horizon. The article finishes with an assessment of the use of the presented formulation to determine delivery plans in practice.

Fuel deliveries

Specialized tankers are used in the distribution of fuels to gas station networks. The most common transport kits are: tractor units and 20 tonne tanker semi-trailers, loadable with approximately 36 thousand liters of fuel. Different types of fuel are transported in chambers with capacities from 2 thousand up to 9 thousand liters.

The primary method of organizing fuel supply is to deliver all types of fuels with one transportation kit to one gas station. In order to maximize the use of the chamber’s capacity, the sum of the different fuel types supplied in a delivery to one station must be equal to the maximum capacity of the vehicle. The second, but less commonly used approach is the so called „multi-split” route where one transportation kit can deliver fuel to more than one gas station. However in the vast majority of cases, the number of stations serviced in this approach does not exceed 3 and is considerably less than the number of gas stations serviced in the case of LPG distribution.

The key element in CMI deliveries is to determine the KPI that the transport company will provide throughout the service life cycle. The main KPI is the availability of fuels at gas stations. In
the vast majority of contracts any situation where stations lack fuel can not take place, meaning that the carrier is obliged to maintain 100% availability of fuel at stations throughout the cooperation period. Fuel shortages often result in high fines for transport companies.

The second KPI group deals with the suggested levels of fuel at gas stations. Safety stock levels form the primary parameter in this group. The transport company should maintain a level not lower than the established safety reserve. The next parameter in this group is the suggested target level at the station. From the gas station’s point of view, this is to limit the stock of fuel in order to limit operation costs. From the point of view of the transport company, the suggested levels are disadvantageous because it limits tank capacity usage at gas stations and causes an increase in the transportation workload. An indirect solution, which on the one hand will limit the fuel level at gas stations, and on the other let a transport company make better use of tank capacity is to apply a suggested level of fuel to the entire network instead of on individual gas stations. This KPI is gaining in popularity yet is very difficult to account for in a manual planning process. However, as presented in the paper, the MIP approach allows us to take this into consideration. As long as the lack of fuel alongside excess capacity are treated as hard constraints, the suggested stocks are used as additional during the planning process and their violation does not have the same effect as a „drying” gas station.

The KPI discussed last is a service level indicator such as the frequency of deliveries over a specified period, interval between deliveries, or days when delivery can not or should be carried out. However, due to the large obstacles that these requirements bring to the planning process, and considering that the vast majority of supplies are delivered to three gas stations at most, KPIs from this group are not often used in practice.

**Tactical planning - planning horizon extension**

One basic way to create fuel delivery plans in accordance with the CMI concept is to analyse current information concerning fuel levels at gas stations and plan deliveries for the next day. The planner must take into account the volume of deliveries on the way and the anticipated demand for each type of fuel at each gas station when preparing the plan. Unfortunately, this mode of operation only helps to prevent fuel shortages at stations, but does not allow for the fulfilment of other KPIs. Such a significant shortening of the planning horizon also fails to allow for limiting the number of kilometers travelled as well as number of transportation kits used.

Due to the above disadvantages of short-term planning (in practice, planning for the next period only) transport companies are trying to extend the planning horizon so they can take advantage of the opportunities offered by CMI and, on the other, prove to distribution companies that achieving appropriate levels of KPIs is guaranteed. This approach results in two main stages of planning. In order to meet these demands, planners, along with ongoing work related to the current planning process, must also plan a weekly supply plan for the entire network of gas stations. However, these activities are very time-consuming, e.g. for a network of stations consisting of 50 stations and 5 types of fuel, it is necessary to consider 1750 potential quantities and delivery dates.

In summarizing the above considerations in the process of planning supply to a gas station network, two basic types of planning can be distinguished: operational and tactical. Operational planning involves assigning specific delivery kits on a given day and the order of delivery taking into account the stock levels obtained from telemetry systems. This planning should also take into account current changes in plans in order to respond to unforeseen situations. Tactical planning is about determining the fuel amount and delivery dates for each gas station, in order to achieve the required levels of relevant KPIs and minimize the transportation workload.

**Literature review**
In operations research, the problems of simultaneous determination of delivery dates and quantities, as well as routes, are known as inventory routing problems (IRPs). The paper by Beltrami and Bodin (1974) may be deemed a pioneering work regarding IRP problems. This work, presented in the 1970’s, focused on modelling and simple solution techniques. IRP research directions can be divided into three main streams.

In the first stream, vehicle routing formulation is extended to take into consideration time horizon and inventory issues. The most promising formulation in this field was proposed by Archetti et al. (2007) and Archetti et al. (2009).

In the second stream of research, the planning horizon is shortened through the computation of suggested replenishment quantities over a long time horizon. Subsequently, a supply timetable and routes for the next few days are determined based on results obtained. For an example in this stream we can refer to a paper by Bard et al. (1998), workload with a rolling horizon of IRP where a short term planning problem is defined for a two-week period. According to the rolling horizon approach, only decisions for the first week are implemented.

The third research stream considers the division of a customer base into delivery groups based on their respective demands and other method-specific parameters. Then, each delivery is performed to all customers in a given group with routes determined with use of the classical VRP or TSP algorithms.

Despite an extended literature review, only a limited number of papers considering a multi-item inventory routing problem were found. Sombat et al. (2005) adopted the policy in which each vehicle always collects the same set of items. A set partitioning problem was implemented to minimize the long-run average inventory and transportation costs. Shan-Huen and Pei-Chun (2010) use the ant colony optimization to plan deliveries to vending machines.

**MIP formulation**

The research method presented is an attempt to use time-discretized integer programming models to solve a real life multi-item IRP. Given the set of problems motivated by real life systems observations (up to 19092 locations, 4 items and a 7-day planning horizon), integer programming has been used to determine fuel distribution. In order to find the solution to fuel distribution problems with less-than-truckload deliveries, a new column generation algorithm proposal was introduced. In comparison to the classical set partitioning formulation, in this approach more than one item and more than one depot are considered. Moreover, based on solutions for full-truck-load deliveries, the column generation approach was proposed. The required stock levels are taken into consideration also.

The main part of the paper contains the MIP model and column generation technique, which allows us to plan the amount and delivery date to the gas station network. The plan does not cover operational issues such as the individual chamber capacity of the kits or the order of deliveries over a period. This information is included during creating operational plans. The tactical plan is superior to the operational plan and is the starting point for its design. However, given the high volatility of fuel demand at gas stations and many additional factors that need to be addressed at the daily planning stage, detailed delivery plans must be modified at the operational stage.

**Computational experiments**

The presented model was used to solve test instances based on data collected from several gas stations in Poland and abroad. Four of instances were examined:

- N65-13 with 65 gas stations and 13 refineries and it corresponds to a real life problem and was a source of motivation for the presented research;
- N106-2 with 106 gas stations and 2 refineries and it corresponds to small gas station networks operated in Poland such as Lukoil or Moya;
- N264-5 with 264 gas stations and 5 refineries, which represents medium size gas station networks like BP;
N1942-33 with 1942 gas stations and 33 refineries, which represents the biggest gas station network in Poland.

In order to find solutions to the presented test instances, three model version were applied. In the first one, limited stock level on each gas station and only FTL deliveries are allowed (denoted as FTL), in the second, the network stock level was considered instead of the stock level on each gas station. Of course the stock level on each gas station was limited to the physical capacity. The second variant is denoted as FTL with network stock. In the final variant, LTL deliveries are accepted and the new column generation approach was applied. Variant denoted as LTL in table 1.

All calculations were made on an Intel i5 2.5MHz computer with the CBC MILP Solver 2.9.7 optimizer running in a Linux environment. In order to limit the solution time, the mixed integer gap was set at 2%. Two planning horizons and five variants of the suggested fuel level on individual gas stations are considered. For instances N65-13-7, N106-2-7 and N264-5-7 the planning horizon was equal to 7 days, while for instance N1942-33-3 the planning horizon was equal to 3 days.

In the context of the suggested stock, it was assumed that the suggested stock on individual gas stations was equal to physical capacity (suggested stock equal to 1), and then the value was decreased to the value of 0.5, by 0.1 in each alternative. Result characteristics (route length, maximal number of kilometres per day, average stock level per product and gas station and the time to find the solution) are shown in Table 1.

Table 1.Key characteristics of solutions to the fuel delivery problem generated on test instances.

<table>
<thead>
<tr>
<th>Instance</th>
<th>sug. stock</th>
<th>FTL</th>
<th>FTL with network stock</th>
<th>LTL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>length</td>
<td>max per day</td>
<td>avg stock</td>
<td>CPU</td>
</tr>
<tr>
<td>N65-13-7</td>
<td>1.0</td>
<td>14210</td>
<td>3956</td>
<td>15366</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>14266</td>
<td>3570</td>
<td>14877</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>14828</td>
<td>3486</td>
<td>14760</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>N106-2-7</td>
<td>1.0</td>
<td>108348</td>
<td>47348</td>
<td>15948</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>109718</td>
<td>45958</td>
<td>15388</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>114228</td>
<td>42998</td>
<td>15134</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>115626</td>
<td>42482</td>
<td>14274</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>N264-5-7</td>
<td>1.0</td>
<td>69054</td>
<td>18298</td>
<td>16767</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>69964</td>
<td>16222</td>
<td>16471</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>71776</td>
<td>29378</td>
<td>15929</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>74948</td>
<td>21224</td>
<td>15182</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>80650</td>
<td>22606</td>
<td>15070</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>N1942-33-3</td>
<td>1.0</td>
<td>70444</td>
<td>30886</td>
<td>20919</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>70444</td>
<td>27508</td>
<td>20473</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>70444</td>
<td>26678</td>
<td>19999</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>73120</td>
<td>29998</td>
<td>19128</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>80650</td>
<td>22606</td>
<td>15070</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

The most important observations are as follows: in many cases with a low value of suggested stock, the transportation company will be not able to plan routes according to FTL.
deliveries. Fuel capacity limitation with FTL routes results in problem infeasibilities (denoted as n/a in the table 1). When the suggested inventory level is low the transportation company has to consider LTL deliveries or network stock. As we can see based on the suggested stock calculation a delivery plan can be provided even for FTL routes and suggested stock level equal to 0.5 of physical capacity. As predicted, the introduction of a suggested stock lowers the average stock of fuel at gas stations in each planning model. The allowance of LTL routes contributed significantly to the achieved values of the solution characteristics. Unfortunately, in the case of FTL supplies, the transport company will not be able to guarantee a suggested stock reduction. On the other hand, the time needed to generate a solution is significantly longer.

Assuming that daily a vehicle can complete only a specified number of kilometres, the maximum number of kilometres per day reported in table 1, allows evaluation of vehicle numbers needed to carry out deliveries (i.e. involved in contract realization). The model with network stock enables a significant decrease in values when compared with FTL solutions.

The results, however, show that the constraints based on a suggested level in the entire network let us reduce the number of kilometres in comparison with FTL solutions and to fulfil customer needs (i.e. a decrease of stock levels). The relationships between the length of the routes and suggested stock level is very important from the point of view of the transport company delivering fuel to gas stations in the CMI approach. The suggested stock level at the individual gas station decreases the average stock level, while it significantly limits the benefit of the CMI concept.

References


Abstract

A motor convoy is a chain of vehicles that are organized together for the purpose of control and secure movement that is dispatched from a source or supplier node to a specific destination in order to provide ammunition, food, medicines, clothing and other requested shipments for demand point.

If there are several suppliers available to satisfy the demands, the configuration of convoys may vary from few long convoys to many small convoys. The Convoy Formation Problem (CFP) is presented to study the best possible configuration of convoys. CFP is a decision about the assignment of vehicles to a convoy and the quantity of shipments delivered by it to the destination. The consequence of this decision not only affects main features of that particular convoy, such as its length and route choice, but also may impact on the formation and movement of other possible convoys on the network. In addition to that, each convoy needs special resources, escorts for protection and medical/maintenance vehicle for emergency situation, limited availability of these resources is another concern which is addressed in CFP.

The output of CFP will be used as the input to the Convoy Movement Problem where the objective function may varies from minimizing the time that it takes to deliver shipments(Chardaire et al. 2005) to finding a route option for convoy which will cause least civilian traffic disruption(Sadeghnejad-Barkousaraie et al. 2016). The main limitation on convoy movement is that two convoys should not meet each other at any point on the transportation network, violation of this constraints is not acceptable, so in this problem we try to find a convoy formation that will minimize the chance of infeasibility of the final solution.

The formation of military convoy is firstly studied by Montana et al. (1999). They studied convoy scheduling, routing and formation problem for a single Origin-Destination pair. Then Robinson and Leiss (2006) used agent based simulation to construct the hierarchy of a convoy formation units e.g. vehicle, columns, serial and march units, and then used genetic algorithm to solve the convoy scheduling problem to find a conflict free solution. In a study by Akgin and Tansel (2007) the deployment planning of military items problem was presented in multi-modal transportation network. In this problem a planner needs to determine the routes, schedule, and assignment of transportation assets in a way that all materials and deploying units arrive at their destinations by their due dates with the goal of minimizing the cost, but formation of convoys has not been addressed there.

In this work we are considering flexibility of choosing suppliers and mandatory formation of convoys, where vehicles can only move in the form of convoys. Where there are O number of suppliers or sources for convoys, S number of shipment and D number of demand points. Each supplier has limited capacity of each shipment which is at most C_s and limited number of vehicles shown as V_o, each vehicle has its own length L_v. There are three different types of vehicles, regular vehicles used for transporting shipments (R) which have the capacity of N_r, maintenance-medical vehicles (M) and escort vehicles (E). The number of regular, maintenance-medical and escort vehicles available in any origin o is limited by R_o, M_o and E_o. Maintenance-medical and escort vehicles are mandatory in formation of convoys. One maintenance-medical vehicle is enough for each convoy with any size and formation, but the number of escort vehicles is relative to the number of vehicles in a convoy and the coverage level of each escort vehicle (H_e). By coverage level we mean the number of vehicles that can be protected by each particular escort vehicle. There are minimum (Min_L) and maximum limits (Max_L) for the length of each convoy and the length of any formed convoy should lie in this range. Also each convoy can have only one origin and one destination and at most one convoy can exist for each pair of origin and destination (od pair). Each vehicle can be assigned to at most one convoy and should not travel by its own, and demands for each shipment in each demand point (I_s) should be fully satisfied. The mathematical representation is provided with model (P).

\[
\begin{align*}
(P) \quad & \text{min} & & w \\
& \text{s.t.} & & \sum_{r \in R_o} \sum_{d} x_{r,s}^d \leq C_s^o \quad \forall s, o \\
& & & \sum_{r} x_{r,s}^d \geq I_s^d \quad \forall s, d \\
& & & \sum_{d} \sum_{s} x_{r,s}^d \leq N_r \quad \forall r
\end{align*}
\]
Now included in the CFP. Instead we assume routes between each origin and destination are known a priori. To avoid adding complexity to this problem, detailed routing constraints are not value the routing consideration is necessary, but the study of convoy routing is NP-Hard problem by itself. To measure this variable which is one when vehicle \( v \) is assigned to a convoy going to destination \( d \). And finally \( w \) is a decision variable representing possible violation of convoys from routing constraints. To measure this value the routing consideration is necessary, but the study of convoy routing is NP-Hard problem by itself (Chardaire et al. 2005). To avoid adding complexity to this problem, detailed routing constraints are not included in the CFP. Instead we assume routes between each origin and destination are known a priori. Now \( w \) is defined as the maximum violation from routing constraints by constraint (10), where \( A_{o,d,o,d'} \) is a parameter representing maximum possible length of convoy from \( o \) to \( d \) if there is a convoy assigned to \( o',d' \), its calculation is presented in algorithm 1.

### Algorithm 1 Calculate Non-conflict maximum length (\( A_{odk,o'd'k'} \))

**Input:** Set \( B \) of all common nodes between path \( od \) and \( o'd' \), average speed of vehicles in \( o \) and \( o' \) as \( VL_o \) and \( VL_{o'} \)

**Output:** The value of \( A_{odk,o'd'k'} \)

1: \( A_{odk,o'd'k'} = MaxL \)
2: for each potential node \( B_i \) in set \( B \) do
3: Calculate \( \hat{T}_{B_i} \), the estimated arrival time to node \( B_i \) for path from \( o \) to \( d \) with start time \( k \)
4: Calculate \( \hat{T}_{B_i} \), the estimated arrival time to node \( B_i \) for path from \( o' \) to \( d' \) with start time \( k' \)
5: Calculate \( \hat{T}_{B_i} = \hat{T}_{B_i} - \hat{T}_{B_i} \)
6: if \( \hat{T}_{B_i} \geq 0 \) then
7: \( L_{temp} = \hat{T}_{B_i} \) \( * VL_o \)
8: if \( A_{odk,o'd'k'} \geq L_{temp} \) then
9: \( A_{odk,o'd'k'} = L_{temp} \)
10: end if
11: end if
12: end for
13: return \( A_{odk,o'd'k'} \)

In order to find a high quality feasible solution in reasonable amount of time we propose a heuristic based decomposition approach, because using commercial solvers such as CPLEX to solve the mathematical model was not successful in this mission. The proposed approach has two major steps. In the first step a feasible solution for the problem is generated, and in the second step the quality of this feasible solution is increased.

A feasible solution of CFP is a set of convoys originated from suppliers ending in demand points which are constituted of transporter vehicles, medical/maintenance vehicles and a set of escort vehicles which are capable of protecting convoy. The amount of shipments transferring by each convoy should not violate the capacity constraints in supply point and should be able to satisfy the demand in demand point. It means there are three main questions which need to be answered, first if there is a convoy from each supplier to demand point, then if there is a convoy, what are the vehicles assigned to it, which their total length (including safety gap between each vehicle in a convoy) lies between specific range. And finally how much of each shipment is assigned to each transporter vehicles of that convoy.

\[
\mathcal{M} y_{d,r} - \sum_s x_{r,s}^d \geq 0 \quad \forall d, r \quad (5)
\]

\[
\mathcal{M} \sum_{m \in M_o} y_{d,m} - \sum_{r \in R_o} y_{d,r} \geq 0 \quad \forall o, d \quad (6)
\]

\[
\sum_{e \in E_o} H_e y_{d,e} - \sum_{v \in V_o, M_o} y_{d,v} \geq 0 \quad \forall o, d \quad (7)
\]

\[
\sum_d y_{d,v} \leq 1 \quad \forall v \quad (8)
\]

\[
z_{o,d} \text{MinL} \leq \sum_{v \in V_o} L_v y_{d,v} \leq z_{o,d} \text{MaxL} \quad \forall o, d \quad (9)
\]

\[
w \geq \sum_{v \in V_o} L_v y_{d,v} - A_{o,d,o',d} + \mathcal{M} (Z_{o',d'} - 1) \quad \forall o, o', d, d' \quad (10)
\]

where \( x_{r,s}^d \) is the amount of shipment \( s \) assigned to vehicle \( r \) to go to demand point \( d \), \( Z_{o,d} \) is a binary variable which equals one when there is a convoy from \( o \) to \( d \) and zero otherwise, \( y_{d,v} \) is another binary variable which is one when vehicle \( v \) is assigned to a convoy going to destination \( d \). And finally \( w \) is a decision variable representing possible violation of convoys from routing constraints. To measure this value the routing consideration is necessary, but the study of convoy routing is NP-Hard problem by itself (Chardaire et al. 2005). To avoid adding complexity to this problem, detailed routing constraints are not included in the CFP. Instead we assume routes between each origin and destination are known a priori. Now \( w \) is defined as the maximum violation from routing constraints by constraint (10), where \( A_{o,d,o',d'} \) is a parameter representing maximum possible length of convoy from \( o \) to \( d \) if there is a convoy assigned to \( o',d' \), its calculation is presented in algorithm 1. 

In order to find a high quality feasible solution in reasonable amount of time we propose a heuristic based decomposition approach, because using commercial solvers such as CPLEX to solve the mathematical model was not successful in this mission. The proposed approach has two major steps. In the first step a feasible solution for the problem is generated, and in the second step the quality of this feasible solution is increased.
Table 1: Indices, Parameters and Decision Variables of Formation Problem.

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Parameters</th>
<th>Decision Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1. ( X_{o,d}^{s,n} ) The amount of shipment ( s ) that is transferred by convoy plan ( n ) from ( o ) to ( d )</td>
<td>( w ) The maximum conflict for all pair of convoys</td>
</tr>
<tr>
<td></td>
<td>2. ( Y_{v,n}^{o,d} ) ( \begin{cases} 1 &amp; \text{if vehicle } v \text{ is used in plan } n \text{ of convoy from } o \text{ to } d \ 0 &amp; \text{otherwise} \end{cases} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. ( Z_{o,d}^n ) ( \begin{cases} 1 &amp; \text{if there is a convoy from } o \text{ to } d \ 0 &amp; \text{otherwise} \end{cases} )</td>
<td></td>
</tr>
</tbody>
</table>

To generate initial solution, the problem is divided to two sections to address these questions sequentially. In the first step, the vehicle assignment is relaxed to find an answer for the first and third questions. Then to find a feasible assignment of vehicles to these convoys, an assignment problem is being solved for each origin separately. If the final solution is not feasible, by using feasible induction a set of constraints are added to both models and then the algorithm will run again. This will continue till a feasible solution is found or infeasibility of the problem is concluded.

Algorithm 2 Recalculating demand satisfaction parameters

1. for each shipment \( s \) do
2. \( TotalCapacity_s = 0 \) and \( counter_s = 0 \) and \( EstimatedLength_s = 0 \)
3. for each origin \( o \) do
4. for each regular vehicle \( r \in R_o \) do
5. \( TotalCapacity_s = TotalCapacity_s + N_{r,s} \)
6. \( EstimatedLength_s = EstimatedLength_s + L_r \)
7. if \( N_{r,s} > 0 \) then
8. \( counter_s = counter_s + 1 \)
9. end if
10. end for
11. end for
12. if \( counter_s > 0 \) then
13. \( AvgCap_s = TotalCapacity_s/counter_s \)
14. \( EstimatedLength_s = EstimatedLength_s/counter_s \)
15. else \( AvgCap_s = 0 \), Stop Problem is infeasible
16. end if
17. for each origin \( o \) do
18. \( CR^o_s = C^o_s/AvgCap_s \), \( CR^o_s \) is vehicle based capacity of shipment \( s \) in origin \( o \)
19. \( L^o_s = EstimatedLength_s \)
20. end for
21. for each demand point \( d \) do
22. \( IR^d_s = I^d_s/AvgCap_s \), \( IR^d_s \) is vehicle based demand of shipment \( s \) in demand point \( d \)
23. end for
24. end for

In the feasible induction section each soft constraint of the vehicle assignment model is studied separately, by relaxing the upper or lower bounds of that constraints. Soft constraints are those which are related to the demand satisfaction plan generated in the first steps, note that the availability, length and the capacity of transporter vehicles can cause the infeasibility in this section. Then constraints that can reflect these limitations are added to the first step or demand satisfaction plan. In order to reflect vehicles characteristics into the demand satisfaction plan, a parameter recalculation is also provided in algorithm 2.

In order to enhance the quality of initial solution, a decomposition method is proposed. In this approach problem is divided into master and sub-problems. Master problem is used to combine solutions found by solving sub-problems with the output of a feasible set of these solutions and in sub-problems actual feasible convoys are generated for each demand point separately and in parallel.

In order to increase the chance of feasibility of a comprehensive solution, a sequence of sub-problems are
being solved where its parameters is updated based on the solution of the previous step. All the solutions provided for each demand points act as columns in restricted master problem in which objective function coefficient of variables in sub-problem are calculated based on the dual values of optimal solution resulted from solving linear relaxation of master problem. Master problem is represented by model \((\mathcal{M}_P)\) and a list of its parameters and decision variables is shown in table 1.

\[
(\mathcal{M}_P) \min \ w \\
\text{s.t. } \sum_n f^n_d = 1 \quad \forall d \\
\sum_d \sum_n X_{o,d}^{s,n} f^n_d \leq C^o_s \quad \forall s, o \\
\sum_d \sum_n Y_{o,d}^{v,n} f^n_d \leq 1 \quad \forall o, v \in V_o \\
w \geq \sum_n \sum_{v \in V_o} (L_v Y_{o,d}^{v,n}) f^n_d - A_{o,d,o'} d' + M(\sum_{n'} \sum_{d,d'} Z_{o,d'}^{n',n} f^n_{d'} - 1) \quad \forall o, o', d, d' \\
f^n_d \in \{0, 1\}, w \in \mathbb{R}_{\geq 0}
\]

The objective function of sub-problem is derived from the dual constraint \(\alpha + \sum_o \sum_s \beta_{o,s} X_{o,d}^s + \sum_o \sum_{v \in V_o} (\gamma_{o,v} - \sum_{o',d'} L_v \lambda_{o,d}^{o',d'}) Y_{o,d}^v \leq 0\) of master problem for each demand point, where \(\alpha, \beta, \gamma\) and \(\lambda\) represent dual variables associated with constraints in \((\mathcal{M}_P)\) respectively.

The expression \(\sum_o \sum_{o'} \sum_{d'} M \lambda_{o,d}^{o',d'}\) is discarded from this dual constraint because the primal constraint (15) associated with this expression, can be active \((\chi_{o,d}^{o',d'} \geq 0)\), only if \(\sum_{n'} Z_{o,d'}^{n',n} f^n_{d'} = 1\) and \(\sum_n Z_{o,d}^{n} f^n_d = 1\) for two specific pairs of \(o, d\) and \(o', d'\).

If \(\sum_{n'} Z_{o,d'}^{n',n} f^n_{d'} < 1\) the right hand side of the constraint (15) should be strictly negative and we already know that \(w \geq 0\) which makes this constraint inactive and therefore its corresponding dual variable \(\chi_{o,d}^{o',d'}\) will be zero\(^2\). We can rewrite constraint (15) as follows:

Constraint (15) : 
\[
\begin{align*}
(\mathcal{M}_B) & \max \sum_o \sum_{v \in V_o} (\gamma_{o,v} - \pi_v L_v) y_v + \omega \\
\text{s.t. } \sum_{v \in R_o} y_v - \sum_{e \in E_o} N_{e} y_e & \leq 0 \quad \forall o \\
\sum_m y_m & \leq \sum_{v \in V_o} L_v y_v \leq \sum_{m \in M_o} y_m \quad \forall o
\end{align*}
\]

\(S_B\) : 
\[
\omega = \max \sum_o \sum_s \beta_{o,s} (\sum_{r \in R_o} x_{r,s}) \\
\text{s.t. } \sum_s x_{r,s} & \leq N_{r} y_r \quad \forall o, r \in R_o \\
\sum_o \sum_s x_{r,s} & \geq I_s \quad \forall s
\]

\(^1\)In case of degeneracy there might be more than two pairs, but it will not change the main concept here

\(^2\)when \(\sum_{n'} Z_{o,d'}^{n',n} f^n_{d'} = 1\) the value of \(\chi_{o,d}^{o',d'} \geq 0\), so it may or may not be zero
To solve sub-problem a Benders’ decomposition approach. The Benders’ representation of subproblem \((M_B)\) is provided as follows where optimality cuts as constraint set \((23)\) where \(\theta'\) and \(\sigma'\) are extreme points of \((S_B)\) and feasibility cuts as constraint \((24)\) where \(\theta\) and \(\sigma\) representing extreme directions of problem \((S_B)\) are added to the model. Flowchart of the proposed algorithm is presented in 1

\[
(M'_B) \quad \max \eta \\
s.t. \quad \eta \leq \sum_{o} \sum_{r \in R_o} (N_r y_r) \theta'_{o,r} + \sum_{s} I_s \sigma'_{s} + \sum_{o} \sum_{v \in V_o} (\gamma_{o,v} - \pi_{o,d} L_v) y_v \\
\sum_{o} \sum_{r \in R_o} (N_r y_r) \theta_{o,r} + \sum_{s} I_s \sigma_{s} \geq 0 \\
\sum_{v \in V_o} y_v - \sum_{e \in E_v} H_v y_e \leq 0 \quad \forall o \\
\text{Min} L \sum_{m \in M_o} y_m \leq \sum_{v \in V_o} L_v y_v \leq \text{Max} L \sum_{m \in M_o} y_m \quad \forall o \\
y_r \in \{0,1\} \quad \forall o, r \in R_o
\]

Finding initial feasible solution

Solving decomposition based approach

Solve demand satisfaction plan
For each origin find a vehicle assignment
Perform Feasibility induction
Is it feasible
Yes
No
Solve master problem \((M_p)\)
For each demand point
Solve Benders decomposition
Update parameters in sub-problem
Add solution as column to master problem
Is it optimal
Yes
No
Add optimality cut
Add feasibility cut
Solve master problem \((M'_p)\)
Does it converge
Yes
No
Solve MIP of Converted problem

Figure 1: Flowchart of the proposed algorithm.

References


Supply Chain Logistics & Methods
FD4: Risk Analysis in Network Transportation
Friday 4:30 – 6:00 PM
Session Chair: Laura Wynter

4:30  Hazardous-Materials Network Design Problem with Behavioral Conditional Value-at-Risk
Liu Su*, Changhyun Kwon
University of South Florida

5:00  A Value-at-Risk (VaR)/Conditional Value-at-Risk (CVaR) Approach to Optimal Train Configuration and
Routing of Hazmat Shipments
S. Davod Hosseini*, Manish Verma
McMaster University

5:30  Entity Resolution and Vessel Modeling for Maritime Situational Awareness
Shiau Hong Lim*, Yeow Khiang Chia, Laura Wynter
IBM Research
Hazardous-materials Network Design Problem with Behavioral Conditional Value-at-Risk

Liu Su and Changhyun Kwon*

Department of Industrial and Management Systems Engineering, University of South Florida

January 4, 2017

1 Introduction

It is an important task to oversee the safe movement for about 1 million daily shipments of hazardous materials (hazmat) crisscrossing the United States. In the past decade, there has been 166,968 incidents causing 105 fatalities, resulting in 2,129 injuries and costing $820,432,788 damage (Pipeline and Hazardous Materials Safety Administration, 2016).

In this paper, we consider a network design problem to minimize the risk of hazmat accidents by selecting a set of road segments to be closed so that hazmat trucks cannot travel, i.e., road bans for hazmat traffic, as considered by Erkut and Gzara (2008), Kara and Verter (2004), and Sun et al. (2016). Typically, network design problems are formulated as a bilevel framework with the upper level to perform a design for decision makers in a network system and lower level to model routing behavior for trucks. However, some existing studies on network design for hazmat incorporates with simplistic methods for routing despite various advanced route choice models. The random utility model (RUM) is a statistical and econometric approach to model route choice which is widely applied to transportation area, for example: Daganzo and Sheffi (1977), Cascetta et al. (1996), Ben-Akiva and Bierlaire (1999) and Ramming (2001).

While modeling probabilistic route-choice of hazmat carriers by RUM, we consider an adverse risk measure called the conditional value-at-risk (CVaR), instead of the widely used expected risk measure. Using RUM and CVaR, we quantify the risk of having hazmat accidents and large consequences, and design the network policy for road bans accordingly. While CVaR has been used in determining a route for hazmat transportation (Toumazis et al., 2013), it has not been considered in the context of route-choice in hazmat network design problems. By applying CVaR to the route-choice behavior of hazmat carriers, we protect the road network from undesirable route-choices that may lead to severe consequences.

*Corresponding Author: chkwon@usf.edu
The proposed model is a nonlinear programming problem which could be decomposed into two stage: (1) searching optimal solution for a single-dimensional variable $r$ to compute CVaR; (2) given each $r$, solving a mixed integer linear programming (MILP) problem involving design variables. A modified golden section method for searching $r$ within a narrowed discrete set is developed. Besides, we could obtain lower bounds and upper bounds to eliminate some $r$ values and improve the computational efficiency without solving MILPs.

2 Model Formulation

2.1 Random Utility and Probabilistic Route Choice Models

Random utility model (RUM) is a widely applied approach to model route choices. Statistically, RUM analyzes the probability associated with choices by incorporating the concept of utility in econometric. It assumes that a traveler’s utility comes from two sources—a deterministic (observable) component and a random (unobservable) component. Given a network, the utility $U_{sk}$ of path $k$ for shipment $s \in S$ between certain origin and destination is defined by:

$$U_{sk} = -\theta^s t_{sk} + \xi_{sk}$$

where $t_{sk}$ is the generalized cost of observable attributes, $\theta^s$ is a positive parameter and $\xi_{sk}$ is a random variable. Usually, $t_{sk}$ is travel time. It could be assumed to be additive with respect to link costs.

$$t_{sk} = \sum_{(i,j)\in A} t_{ij} \delta_{ij}$$

where $t_{ij}$ is the generalized travel cost for link $(i,j)$, and $\delta_{ij} = 1$ if link $(i,j)$ is on path $k$ for shipment $s$ and 0 otherwise. To obtain an explicit form for route choice probabilities, it could be assumed that $\xi_{sk}$ are independently and identically distributed from Gumbel variates resulting in Multinomial Logit (MNL) model. According to Sheffi (1985), the probability of choosing path $k$ for shipment $s$ is,

$$\pi_{sk} = \frac{e^{-\theta^s t_{sk}}}{\sum_l e^{-\theta^s t_{sl}}}.$$
number of shipments for shipment $s \in \mathcal{S}$. The risk measure for a transportation network could be:

$$R = \sum_{s \in \mathcal{S}} \sum_{l=1}^{n_s} R_i^s$$

(4)

where $R_i^s$ is the risk for demand $l$ in shipment $s$. Let $C_{(m)}$ denote the $m$-th smallest value in \{ $c_{ij}^s : (i, j) \in \cup_{s \in \mathcal{S}, k \in \mathcal{K}_s} \mathcal{A}^{sk}$ \} with accident probability $p_{(m)}$. In addition we let $c_{ij}^s$ denote the accident consequence with probability $p_{ij}$ for shipment $s$ on link $(i, j)$, and $\mathcal{A}^{sk}$ denote the set of links that path $k$ contains for shipment $s$. We define $M_A$ to be equal to $\sum_{s \in \mathcal{S}} |\cup_{k \in \mathcal{K}_s} \mathcal{A}^{sk}|$. The risk for a hazardous transportation network is defined by,

$$R = \begin{cases} 0 & \text{with probability } 1 - \sum_{s \in \mathcal{S}} \sum_{k \in \mathcal{K}_s} \sum_{(i, j) \in \mathcal{A}^{sk}} n_s \pi^{sk} p_{ij} \\ C_{(1)} & \text{with probability } \sum_{k \in \mathcal{K}_1} \sum_{(i, j) \in \mathcal{A}^{sk}} n_s \pi^{sk} p_{(1)} \\
\vdots & \vdots \\ C_{(M_A)} & \text{with probability } \sum_{k \in \mathcal{K}_{M_A}} \sum_{(i, M_A, j, M_A) \in \mathcal{A}^{sk}} n_s \pi^{sk} p_{(M_A)} . \end{cases}$$

(5)

We define CVaR as an adverse risk measure for the above probabilistic risk variable.

### 2.3 CVaR Minimization Model for Network Design

The CVaR minimization model is defined as follows,

$$\min_{y,z} \text{CVaR}_\alpha = \min_{y,z} \Phi_\alpha(r; y, z)$$

(6)

where $\Phi_\alpha(r; y, z) \triangleq r + \frac{1}{1-\alpha} \mathbb{E}[R - r]^+$ and

$$\Omega = \{ \text{Linearization for RUM} \}$$

(7)

Path-based network design constraints.

Let $y$ and $z$ be the path-based network design variables and $r$ be variable for analyzing CVaR. Note that the random risk variable $R$ depends on $y$ and $z$. Besides, $\Theta = \{0\} \cup \{c_{ij}^s : (i, j) \in \cup_{s \in \mathcal{S}, k \in \mathcal{K}_s} \mathcal{A}^{sk}\}$ based on (5). Route choice probability (3) depends on network design which could be linearized. Let $f_{\alpha}(r) = \min_{y,z} \Phi_\alpha(r; y, z)$ and $r_q$ be the $q$-th smallest value in $\Theta$. Optimal solution for $r$ is searched within $\Theta$. Given $r$, MILPs are solved to obtain optimal solution for the proposed model. In order to eliminate impossible $r$ values and improve the computational efficiency, we analyze $f_{\alpha}(r)$.

**Theorem 1.** For some $q \in \{0, 1, \ldots, M_A\}$, suppose the following condition holds:

$$\frac{1}{1-\alpha} \sum_{(i, j) \in \mathcal{A}} \sum_{s \in \mathcal{S}} \sum_{k \in \mathcal{K}_s} c_{ij}^s \geq r_{q+1} n_s \pi_{ij} \leq 1$$

(9)
Then we can show that
\[
\Phi_\alpha(r_q; y, z) \leq \Phi_\alpha(r_{q+1}; y, z)
\]
for all \(y, z \in \Omega\). Further we can show that \(f_\alpha(r_q) \leq f_\alpha(r_{q+1}) \leq f_\alpha(r_{q+2}) \leq \cdots \leq f_\alpha(r_{M_A}).\)

Theorem 1 is used to expedite the search process in the computational method proposed in the next section.

3 Solution Techniques

The proposed CVaR minimization model can be decomposed into two stages. In the first stage, we search \(r\) within a finite set \(\Theta\) as the distribution of network risk for accident consequence is discrete. In the second stage, a mixed-integer linear programming (MILP) problem is solved to yield a network design solution. We can obtain an optimal solution for network design problem by visiting every value in the set \(\Theta\). If \(\Theta\) involves a large number of values, computation for the problem would be extremely time-consuming because we need to solve a large number of MILPs. A searching mechanism for \(r\) based on the golden section method is proposed in order to solve the problem efficiently. Typically, the golden section method applies to a strictly quasiconvex over an interval. The core for the golden section method is reusing one searching point in previous iteration and comparing to an updating point derived by golden ratio to reduce computations. We develop a discrete version of the golden section method which searches value in \(\Theta\) based on the same idea. Based on Theorem 1, some values in \(\Theta\) could be excluded while searching \(r\) thus reducing computation efforts. Hence, we could incorporate the golden section method with Theorem 1 and proposed a computational scheme for CVaR Minimization model for network design.

The computational efforts for the search process can be further reduced. Instead of solving the MILP problem for each search for \(r\) optimally, we can obtain lower and upper bounds using techniques based on Lagrangian relaxation. Using these bounds, we can eliminate some \(r\) values within relatively short time.

4 Computational Studies

We applied the proposed model using Ravenna Bonvicini and Spadoni (2008) network data transporting four kinds of hazmat: chlorine, LPG, gasoline and methanol. The network has 105 nodes and 134 undirected links. We implemented our search algorithm on the network and the network design results are shown in Figure 1.

References


A Value-at-Risk (VaR)/Conditional Value-at-Risk (CVaR) Approach to Optimal Train Configuration and Routing of Hazmat Shipments

S. Davod Hosseini¹
Dr. Manish Verma²

January 04, 2017

¹ PhD Candidate in Management Science at the DeGroote School of Business, McMaster University, Canada
² Associate Professor in Management Science at the DeGroote School of Business, McMaster University, Canada
Introduction

Hazardous materials (hazmat) are harmful to humans and the environment because of their toxic ingredients, but their transportation is essential to sustain our industrial lifestyle. A significant majority of hazmat shipments are moved via the highway and railroad networks. For example, in the United States, railroad carries approximately 1.8 million carloads of hazmat annually, which translates into 5% of rail freight traffic. On the other hand, in Canada, approximately 500,000 carloads of hazmat-equivalent to 12% of total traffic—are shipped by railroad. It may appear that railroads are not the predominant mode for surface transportation of hazmat, but they are almost always preferred to move shipments over long distances. In the United States, for instance, railroads account for around 29% of hazmat movement in ton-miles compared to 32.2% for trucks, which in turn translated into a 27.9% increase from 2002 to 2014.

The quantity of hazmat traffic on the railroad network is expected to increase significantly over the next decade, given the phenomenal growth of intermodal transportation and the growing use of rail-truck combinations to move chemicals. It is true that railroads have a favorable safety statistic, but the possibility of spectacular events resulting from multiscar incidents, however small, does exist. The derailment of the BNSF train in Lafayette (Louisiana, United States), spilling 10,000 gallons on hydrochloric acid and forcing more than 3,000 residents out of their homes, is an example of such low probability–high consequence events. In fact in the United States, between 1995 and 2009, around 120 train accidents resulted in release from multiple tank cars, which translates into an average of eight accidents every year.

Hazmat accidents rarely happen (low-probability incidents), but if they do occur, then the consequences can be disastrous (high-consequence incidents), reflecting on both the population and the environment. Therefore it is important to make a risk-averse route decision in hazmat transportation. In this study, for the first time we will develop Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) optimization models applied in railroad transportation of hazmat. VaR and CVaR (first introduced in finance as a risk measure) will be shown to be proper risk measures for flexible hazmat route decision making.

Compared to the other existing approaches for hazmat routing, VaR and CVaR produce a more flexible and reliable route modeling approach for hazmat transportation. Depending on the decision makers’ attitude to risk, one can make multiple planning decisions according to each individual risk preferences. Instead of a single optimal route output, these models have a two-dimensional framework which produces alternative route choices given different confidence levels. In addition, while most existing hazmat routing methods study the entire risk distribution, CVaR obviously focuses more on the long tail to avoid extreme events. In fact, being risk-averse by focusing on the long tail is more reasonable for hazmat transportation.

Rail-route Risk

In a railroad transportation system, the physical infrastructure comprises of rail yards and tracks. Any two nodes are connected by tracks, which are the service legs of a train traveling nonstop between them. A sequence of service legs and intermediate yards constitutes a route available to a railcar for its journey. There are limited train services in the network to perform these journeys. The objective here is to transport a specific number (N) of hazmat railcars (along with regular railcars) between single origin-destination (O-D) yards using the available train services in the network. The hazmat railcars are exposed to derailment and then release of hazmat if the train meets with an accident in the tracks and/or at the yards. Therefore, we consider a directed railroad network \( G = (\mathcal{Y}, \mathcal{A}) \), where \( \mathcal{Y} \) shows the set of yards. For each service leg \((i, j) \in \mathcal{A}\) (or yard \(k \in \mathcal{Y}\)), there are two attributes: accident probability denoted by \( p_{ij} \) (or \( p_k \)) and accident consequence denoted by \( c_{ij} \) (or \( c_k \)). In fact, they are the probability that a train meets with an accident on this service leg (or at this yard) and the resulting consequence, respectively. Based on this exposition, we define the following notations:
\[ y: \text{Set of yards, indexed by } i, j, k \]
\[ A: \text{Set of service legs, indexed by } (i, j) \]
\[ S: \text{Set of train services, indexed by } s \text{ (and/or } \hat{s}) \]
\[ N: \text{Number of hazmat railcars to be shipped} \]
\[ p_k: \text{Accident probability at yard } k \]
\[ p_{ij}: \text{Accident probability in the service leg connecting yards } i \text{ and } j \text{ together} \]
\[ c_k: \text{Accident consequence at transferring yard } k \]
\[ c_{ij}: \text{Accident consequence in the service leg connecting yards } i \text{ and } j \text{ together} \]

To be able to assess risk at a high resolution, we compute derailment probabilities based on the ten *deciles* of the train (the length of the train is divided into 10 equal parts). For similar train-lengths, we define the positive integer \( y_r \) as the number of hazmat railcars in decile \( r \) of the train. This way, \( \sum_{r=1}^{10} y_r = N \), where \( y_r \leq \frac{\text{train length}}{10} \).

Utilizing the above notations and explanations, \( p_{ij} \) and \( p_k \) would be determined as:

\[
p_{ij} = P(A_{ij}) \times \sum_{r=1}^{10} y_r \times \left( P(D^r|A_{ij}) \times P(H|D^r,A_{ij}) \times P(R|H,D^r,A_{ij}) \right)
\]

\[
p_k = P(A_k) \times \sum_{r=1}^{10} y_r \times \left( P(D^r|A_k) \times P(H|D^r,A_k) \times P(R|H,D^r,A_k) \right)
\]

where \( P(A_{ij}) \) (or \( P(A_k) \)) is the probability that a train meets with an accident on service leg \((i, j)\) (or at yard \(k\)); \( P(D^r|A_{ij}) \) (or \( P(D^r|A_k) \)) is the probability of derailment of a railcar in the \(r\)th decile of the train given the accident on service leg \((i, j)\) (or at yard \(k\)); \( P(H|D^r,A_{ij}) \) (or \( P(H|D^r,A_k) \)) is the probability that a hazmat railcar derailed in the \(r\)th decile of the train given the accident on service leg \((i, j)\) (or at yard \(k\)). In addition, \( c_{ij} \) and \( c_k \) would be determined as the population exposure (PE) due to the release from \(N\) hazmat railcars given the accident on service leg \((i, j)\) and at yard \(k\), respectively. Train-accident rates are estimated using the Federal Railroad Administration (FRA) website, while the conditional probabilities are estimated by applying logical diagrams and Bayes Theorem to the dataset (FRA, 2010) containing derailment statistics from 1995 to 2009.

**Value-at-Risk (VaR) Minimization Model**

Let \( R^l \) denote the discrete random variable for the risk associated with route \( l \). The \( \alpha \)-VaR value for the risk random variable associated with route \( l \) and any specified confidence level \( \alpha \) in \((0,1)\) will be denoted by \( \text{VaR}^l_\alpha \) and is given by

\[
\text{VaR}^l_\alpha = \min \{ \beta | \Pr(R^l \leq \beta) \geq \alpha \}
\]

which means that VaR is the minimal level \( \beta \) such that the risk measure \( R^l \) does not exceed \( \beta \) with the least probability of \( \alpha \). After doing a lot of computation, we reach the following final model

\[
\min_r \mathcal{C}_{(r)}
\]

\[
f^r = \min_x \mathcal{P}_{(r)} X
\]

subject to
\[ f^r \leq 1 - \alpha < f^r + P(r) \]
\[ X \in \psi \]
\[ r \in \{0,1,2,\ldots,M\}. \]

The solution of this model (using an efficient shortest path algorithm like Dijkstra’s Algorithm with some modifications) produces VaR\(^\star\) and its corresponding route for the shipment of the \(N\)-hazmat railcars from the origin to the destination in the network. \(C_{(r)}\) is the sorted accident consequences of all \((M)\) yards and service legs in the network and \(P_{(r)}\) is their corresponding accident probabilities, \( r \in \{0,1,2,\ldots,M\}\). \( \mathbb{P}_{(r)} \) and \( X \in \psi \) (for train routing in the network) are defined as follows

\[
\mathbb{P}_{(r)} = \sum_{s, s'} \left( \sum_{k \in y_s} \sum_{s''} p_{k,s'} + \sum_{s''} p_{i_s,j_s} \right), \quad r \in \{0,1,2,\ldots,M\} 
\]

\[
x_{i_s,j_s} = \begin{cases} 
1, & \text{if service leg } (i_s,j_s) \text{ from train service } s \text{ is used in the route} \\
0, & \text{otherwise} 
\end{cases}
\]

\[
x_{k,s} = \begin{cases} 
1, & \text{if yard } k \text{ is used in the route as a transferring yard between train services } s \text{ and } \hat{s} \\
0, & \text{otherwise} 
\end{cases}
\]

\[ \sum_{j_s} x_{i_s,j_s} = 1 \quad \text{for origin yard } i \text{ and train service } s \]

\[ \sum_{j_s} x_{j_s,i_s} - \sum_{j_s} x_{i_s,j_s} = 0 \quad \text{for any non-transferring yard } i \text{ and train service } s \]

\[ \sum_{j_s} x_{j_s,k_s} - \sum_{j_s} x_{k_s,j_s} = 0 \quad \text{for any transferring yard } k \text{ and train service } s \text{ and } \hat{s} \]

\[ \sum_{j_s} x_{j_s,i_s} = 1 \quad \text{for destination yard } i \text{ and train service } s. \]

**Conditional Value-at-Risk (CVaR) Minimization Model**

VaR is a relatively simple risk management notion; it has a clear interpretation: how much you may lose with certain confidence level. Despite its wide usage, VaR has been criticized as a risk measure due to the fact that it is not a coherent risk measure and it might lead to inaccurate perception of risk. More specifically, a very serious shortcoming of VaR is that it provides no handle on the extent of the losses that might be suffered beyond the threshold amount indicated by this measure. It is incapable of distinguishing between situations where losses that are worse may be deemed only a little bit worse, and those where they could well be overwhelming.

For continuous loss distributions, the CVaR at a given confidence level is the expected loss given that the loss is greater than the VaR at that level, or for that matter, the expected loss given that the loss is greater than or equal to the VaR. These conditional expectations are designated by CVaR\(^\star\), called “upper CVaR”, and CVaR\(^\circ\), called “lower CVaR”, respectively. Therefore for confidence level \(\alpha\) and random variable \(X\) (representing loss, for example) with continuous distribution function \(F_X(x)\), CVaR\(\alpha\) (\(X\)) equals the conditional expectation of \(X\) subject to \(X \geq \text{VaR}_\alpha(X)\); CVaR\(\alpha\) (\(X\)) = \(E[X \mid X \geq \text{VaR}_\alpha(X)]\). For distributions with possible discontinuities, however,
CVaR has a more subtle definition and can differ from either of CVaR$^+$ and CVaR$^-$ quantities. Therefore, contrary to popular belief, in the general case, $\text{CVaR}_\alpha(X)$ is not equal to an average of outcomes greater than $\text{VaR}_\alpha(X)$. The $\alpha$-CVaR of the risk associated with the route $l$ is defined as the value

$$ \text{CVaR}_\alpha^l = \text{mean of the } \alpha \text{-tail distribution of } R^l, $$

where the distribution in question is the one with distribution function (CDF) $F_{R^l}^\alpha(\beta)$ defined by

$$ F_{R^l}^\alpha(\beta) = \begin{cases} 0, & \text{for } \beta < \text{Var}_\alpha^l \\ \frac{F_R^l(\beta) - \alpha}{1 - \alpha}, & \text{for } \beta \geq \text{Var}_\alpha^l \end{cases} $$

hence

$$ \text{CVaR}_\alpha^l = \int_0^\infty \beta \, dF_{R^l}^\alpha(\beta). $$

**CVaR for Scenario Models**

Consider the route $l$ and recall that the distribution of the risk $R^l$ is concentrated in finitely many points; $C^l_i, i \in \{0, 1, 2, \ldots, n_l\}$, where $C^l_0 < C^l_1 < C^l_2 < \cdots < C^l_{n_l}$ and the probability of $C^l_i$ being $P^l_i$. This way, the distribution function $F_{R^l}(\beta)$ is a step function with jumps at those points. Let $K$ be equal to 0 if $0 < \alpha \leq P^l_0$, otherwise let it be the unique index such that

$$ \sum_{i=0}^{K-1} P^l_i < \alpha \leq \sum_{i=0}^K P^l_i. $$

The $\alpha$-VaR of the risk is given by

$$ \text{VaR}_\alpha^l = C^l_K $$

then we show that the $\alpha$-CVaR is given by

$$ \text{CVaR}_\alpha^l = \frac{1}{1 - \alpha} \left[ \left( \sum_{i=0}^{K} P^l_i - \alpha \right) \text{Var}_\alpha^l + \sum_{i=K+1}^{n_l} P^l_i C^l_i \right]. $$

**CVaR Optimization**

After doing several steps, the $\text{CVaR}_\alpha^l$ minimization problem will be

$$ \text{CVaR}_{\alpha}^* = \min_r \text{CVaR}_\alpha^r $$

subject to

$$ \text{CVaR}_\alpha^r = C_{(r)} + \frac{1}{1 - \alpha} f^r $$

$$ f^r = \min_X \hat{p}_{(r)} X $$

$$ X \in \psi $$

$$ r = 0, 1, 2, \ldots, M $$
where

$$\hat{P}(r) = \sum_{s, s'} \left( \sum_{k, s \in Y_s \& s' \in Y_{s'}} \left( p_{k,s} (c_{k,s} - C(r)) + \sum_{(s, s') \in \mathcal{A}_s} p_{i,j} (c_{i,j} - C(r)) \right) \right),$$

and $f^r$ minimization will be done using an efficient shortest path algorithm like Dijkstra’s Algorithm if we make some modifications.

### Train Configuration Setting

The accident probabilities in the service legs and at yards, i.e. $p_{ij}$ and $p_k$, respectively, depend on determining the number of hazmat railcars in each decile of the train (we call it Train Configuration). There are many different configurations that can lead to different values for risk measures, including VaR and CVaR. Hence we should determine the train configuration in a way that the resulting accident probabilities for both service legs and yards consequently lead to the minimum values of the risk measures and their corresponding optimal routes.

Let define constants $T_{CP}^r$ and $Y_{CP}^r$, $r = \{1,2, \ldots ,10\}$, as the multiplication of conditional probabilities in decile $r$ of the train for service legs and yards, respectively, i.e. $T_{CP}^r = P(D^r|A_{ij}) \times P(H|D^r,A_{ij}) \times P(R|H,D^r,A_{ij})$ and $Y_{CP}^r = P(D^r|A_k) \times P(H|D^r,A_k) \times P(R|H,D^r,A_k)$. Given the above explanation, we define the following minimization model to find the best train configuration:

$$\min_{W,Y_r} \left( \sum_{r=1}^{10} y_r \times T_{CP}^r \right) + (1-W) \left( \sum_{r=1}^{10} y_r \times Y_{CP}^r \right)$$

subject to

$$\min_{l \in \mathcal{P}} \text{VaR}_{\alpha}^l \text{ or } \min_{l \in \mathcal{P}} \text{CVaR}_{\alpha}^l$$

$$\sum_{r=1}^{10} y_r = N$$

$$0 \leq y_r \leq \frac{\text{train length}}{10}$$

$$y_r: \text{integer}, \quad r = \{1,2, \ldots ,10\}$$

$$0 \leq W \leq 1.$$  

### Computational Experiments

We use the proposed methodologies for solving several problem instances to develop managerial insights for railroad transportation of hazmat. We represent the railroad infrastructure in the Midwest United States, which is used to generate realistic-sized problem instances. There are 25 yards in the network. Each node can be both an origin and destination for the others, and hence there are 600 origin-destination pairs. We also consider 31 train services available in this network. The objective is to determine the best way to move $N$ hazmat railcars between various O-D pairs using the available train services in the network such that the transport risk measuring by VaR and CVaR is minimized. Note that this implies not just a decision about the route but also the placement of hazmat railcars in a train. We report on our analyses conducted on the instances generated based on this problem setting.
Entity Resolution and Vessel Modeling for Maritime Situational Awareness
Shiau Hong Lim, Yeow Khiang Chia and Laura Wynter

Abstract
We present methods for two challenging problems in the maritime domain, namely entity resolution and vessel activity modeling.

1 Introduction
Maritime situational awareness concerns the use of various information sources, including both real-time and historical data, to enhance the decision making process in maritime operations. Example activities of importance to maritime operations include the tracking of vessel movements, the verification of reported operations and the automatic detection of infringements. Accurate prediction of maritime vessel movement enhances planning and resource utilization, and allows preemptive actions to be taken to avoid congestion and prevent accidents.

Two key components that provide the foundation for many maritime operations activities are Entity resolution and Vessel Modeling.

2 Entity Resolution
Entity resolution seeks to establish the identity associated with any given piece of information or data record. We assume that various sources of information, including AIS broadcasts and radar tracks, have been processed and fused into a sequence of vessel traffic records. Each record typically contains one or more pieces of information on vessel identity, such as the IMO and the MMSI numbers, region-specific vessel ID, as well as the radar tracking ID. Incomplete and incorrect IDs are prevalent and the challenge is to resolve any confusions quickly and with confidence. Our approach takes as input the individual vessel traffic records and assigns a unique identifier to each record in real time. The identifiers may be updated when further information is received.

The problem of entity resolution can be stated as follows. Given a set of records $R = \{r_1, \ldots, r_n\}$, one seeks a corresponding set of labelings $Q = \{q_1, \ldots, q_n\}$ such that $q_i = q_j$ if and only if $r_i$ and $r_j$ both refer to the same physical vessel. One can build a model that assigns a likelihood score $L(R, Q)$ for any labeling $Q$ of $R$, and solve for $\arg \max_Q L(R, Q)$. In general this is a hard combinatorial problem and the challenge is to find an approximate solution that is computationally tractable in real time.

Our solution is based on the following two models:

1. A model of physical feasibility. Given a set of records $R$, it assigns a score based on how likely the entire set $R$ comes from a single unique vessel. The model needs to strike a balance between robustness to noisy data and sensitivity to violation of physical laws.

2. An model of identity matching. Given two sets of records $R_1$ and $R_2$, each containing some ID information, the model assigns a score based on how likely $R_1$ and $R_2$ both refer to the same vessel.

The first model is used to detect the fact that a set of records may involve more than one vessel and is used to partition the set into physically coherent subsets. The second model is the basis for merging distinct
sets of records that are in fact referring to the same vessel, for example, due to a transfer of ownership and/or a change of MMSI number.

An overview of the approach developed for maritime entity resolution will be covered during the presentation.

3 Vessel Modeling

Both prediction and anomaly detection rely on a model of typical vessel activities, including potential interactions with other vessels. The better the model can predict any future activities, the better it can detect abnormal behaviors. While various approaches have been proposed in this area [1–6], we briefly describe our approach, which is unique in the way it combines both micro and macro-level events.

3.1 Micro events

We consider each new vessel record a micro event. This typically results from a new AIS broadcast that contains the most up-to-date positional and speed information of a vessel. Chaining together such events forms a trajectory in space and time. We take a nonparametric approach to trajectory modeling, and avoid making unnecessary assumptions or model simplifications. For the purpose of predicting the near-term future, the key challenge in this approach is to quickly retrieve the most relevant set of records from a large volume of historical records. Advances in nearest-neighbor lookups make this possible in real time.

To illustrate, Figure 1 shows the errors in predicting the exact location of a vessel in 30 minutes based on the most recent 20 minutes of trajectory records. The trajectory is encoded as a sequence of coordinate pairs, equally spaced in time via linear interpolation. We compare the simple \( k \)-nearest-neighbor approach with the more sophisticated approach based on Gaussian processes (labeled GP). The GP-fused approach will be discussed in the next section. It is interesting to note that the GP approach, based on the same 50 nearest neighbors, performs very similarly to the \( k \)-NN approach that makes use of only simple averages.

3.2 Macro events and Fusion

In order to maximize the prediction accuracy, one could imagine trying to use all of the available information, such as the full historical trajectory of a vessel. Multiple levels of abstraction is almost certainly necessary if one would want to make this practical. For example, one could abstract away data that are less recent and use a coarser scale, in order to retain the conciseness in representations. We refer to these abstractions
over a larger scale in both space and time as macro events. In particular, we partition the historical trajectories of a particular vessel (or class of similar vessels) into segments of MOVE, STOP, and DISAPPEAR, where DISAPPEAR can mean either moving out of range of the port or simply turning off the transponder broadcast. Each segment type is clustered based on physical locations and/or time durations.

We build a Hidden Markov Model (HMM) of macro events, one for each vessel or vessel-class. The HMM, given any sequence of events, can produce a distribution over the likely future macro events. The posterior distribution can be transformed into a feature vector, and combined with the features from micro events to produce an augmented feature vector. The GP can be tuned to optimize the relative weights between the micro and macro events. The predictive results is reported as “GP-fused” in Figure 1. We see a significant improvement in prediction error. Figure 2 shows an example of such improvement. Here, the query trajectory is in red. The nearest-neighbors (in terms of micro events) are all very similar and in this particular case the 1-NN approach performs badly. The 30-minute lookahead in fact consists of two rather distinct destination, but averaging the two will result in a location far from both destinations. The macro-event part of the features helps filtering out the “wrong” group of neighbors.

One immediate application of the fused predictive model is in improving the prediction of whether a particular vessel will pass through a controlled channel. This approach is particularly suitable in predicting the behavior of a large group of vessels, such as tankers, that operate regularly in a particular region or port area. Accurate predictions will help reduce communication overheads and enhance the overall safety, especially in highly utilized ports or regions.

The details of our proposed vessel modeling approach will be covered in the presentation.

References

SUPPLY CHAIN LOGISTICS & METHODS
SA4: BEHAVIORAL DATA AND DEMAND ESTIMATION
Saturday 9:00 – 11:00 AM
Session Chair: Yueyue Fan

9:00  Household Use of Autonomous Vehicles: Modeling Framework and Traveler Adaptation
      Yashar Khayati, Jee Eun Kang*, Mark Karwan, Chase Murray
      University at Buffalo

9:30  Estimating Primary Demand of One-Way Vehicle Sharing Systems
      Chiwei Yan*, Chong Yang Goh
      Massachusetts Institute of Technology

10:00 Modeling the Acceptability of Crowdsourced Goods Deliveries
      Aymeric Punel, Alireza Ermagun, Amanda Stathopoulos*
      Northwestern University

10:30 Travel Demand Estimation Using Heterogeneous Data Pieces: Addressing Stochasticity and Observability Issues
      Yudi Yang, Yueyue Fan*, Roger Wets
      University of California, Davis
Household Use of Autonomous Vehicles: Modeling Framework and Traveler Adaptation

Yashar Khayati, Jee Eun Kang, Mark H. Karwan and Chase C. Murray
Industrial and Systems Engineering
University at Buffalo, State University of New York
January 2nd, 2017

1 Introduction

With recent developments in vehicle automation, the idea and advantages of sharing public roads with Autonomous Vehicles (AV) or self-driving vehicles has been accepted. Many aspects of today’s transportation system will be transformed with the introduction of AVs. These changes are attributed broadly to two properties of AVs: automated control and driverless operations. (Pinjari, 2013), and similarly (Litman, 2014), argue that potential impacts include (1) increased safety, (2) better use of traveler travel time, (3) independent mobility for the elderly and disabled populations, (4) reduced fuel consumption and emissions and (5) increased road capacity. It can be categorized that (1), (4), (5) come from automated control, and (2), (3) come from driverless operations. In the longer term, driverless operations may also affect individuals’ home location choice, vehicle ownership choice, etc., resulting in reshaping of land-use characteristics and the transportation system.

Three areas which are expected to be greatly transformed are vehicle use, travel behavior and subsequently vehicle ownership from driverless operations. Ever since the introduction of personal automobiles in the United States, per capita ownership grew during most of the 20th century and it has been the norm that a licensed driver operates his/her own vehicle. In 2011, there were 0.75 vehicles per person and 1.10 vehicles per licensed driver, and 1.95 vehicles per household in the US. (Sivak, 2013; Litman, 2014). With AVs, this travel behavior is expected to change significantly. It is likely that travelers in a household can share one vehicle with its driverless operations. On the other hand, it is also possible that travelers without a license (e.g. children, disabled or elderly) could own a personal AV. Another prediction is that travelers will utilize autonomous taxis, or shared-mobility systems, as the cost of these systems will be greatly reduced with no driver involved (Pinjari, 2013).

In this paper, we define a framework to model and evaluate potential household-level use of AVs, to understand advantages, potential issues and negative external effects. Since there are no data readily available in this field, we introduce a new formulation; the Household Activity Pattern Problem for Autonomous Vehicles (HAPP-AV) to simulate the travel patterns of people using AVs. The Household Activity Pattern Problem (HAPP) is a constraint-based full-day activity scheduling problem and can be used to simulate potential use of policies or adoption of new vehicles (Recker and Parimi, 1999; Kang and Recker, 2014). Based on the framework of HAPP, HAPP-AV additionally includes the self-driving capabilities of AVs.

*corresponding author, email: jeeunka@buffalo.edu
Using AVs in a household vehicle stable leads to significant differences between HAPP and HAPP-AV. Since no driver needs to be assigned to vehicles, AVs can be more effectively utilized, serving more trips during the day. The key modeling challenge is to include modeling capabilities of ride sharing, driverless operation, parking, pickup and dropoff as well as waiting during travelers’ engagement in activities. Computationally, these new features make AV routing significantly more complicated than the original HAPP. We developed different solution approaches to solve this NP-hard problem. Using this model, we plan to conduct a scenario analysis to draw insights on changes in travel behavior and vehicle use, including vehicle-miles-traveled (VMT) and emissions, among other metrics.

2 The Household Activity Pattern Problem for Autonomous Vehicles

The decision making scheme of the proposed HAPP-AV is as follows. Household members participate in a given set of out-of-home activities using a set of household autonomous vehicles while minimizing associated travel disutilities, within spatial and temporal constraints. HAPP-AV can be divided into two basic routing problems. One problem is to route household members to perform out-of-home activities within their time windows (temporal constraints) and at specific locations (spatial constraints). The second problem is to route AVs to pickup and dropoff household members and also decide the best places to park. These two problems should be integrated and solved simultaneously in order to incorporate routing flexibility of travelers as well as vehicles. A conceptual illustration is given in Figure 1.

HAPP-AV falls into a class of Vehicle Routing Problems (VRP). In particular, HAPP-AV includes many types of synchronization (Drexl, 2012). There are different types of synchronizations in the VRP literature such as task, operation, movement, resource and load synchronization. Most of the VRPs in the current literature only include one or two of the synchronizations; however in HAPP-AV we need to consider task, operation, movement and load synchronizations at the same time makes HAPP-AV computationally challenging. The key modeling challenge lies in making the model as general and as realistic as possible. This model must accommodate numerous cases that can occur in daily operations of AVs; such as parking availability in activity locations, household members returning home during the day, an AV traveling without any passenger or with more than one passenger, etc. Considering all of those features requires us to model multiple synchronizations. Thus, we duplicate some node sets of the same physical locations and introduce four major decision variables to model the routing problem and capture travel behavior of people while using AVs.

General structure of HAPP-AV is as follows.

$H_{ij}^p$ is a binary variable which is one if household member $p$ travels from node $i$ to node $j$.

$X_{ij}^v$ is a binary variable which is one if AV $v$ travels from node $i$ to node $j$.

$T_i$ is the start time of node $i$.

$W_i^p$ is the waiting time of household member $p$ to start or be picked up at node $i$.

$C_i^v$ is the number of passengers in the AV $v$ a moment after arriving to node $i$.

$H_{ij}^p$ and $X_{ij}^v$ are employed to capture travelers and AVs travel decisions respectively. $T_i$ is the time at which node $i$ is visited either by an AV (such as parking) or by a household member and an AV simultaneously (such as pickup and drop-off nodes). $W_i^p$ is used to penalize the long waiting times for household members. Controlling the infeasible flows in the AVs network by explicit constraints is difficult; thus, we used $C_i^v$ variables to avoid infeasible flows. $C_i^v$ can be used to consider some specific personal preferences of travelers such as desiring to travel alone.
The constraints defining HAPP-AV can be grouped into six broad categories (a) temporal constraints on the vehicles, (b) temporal constraints on the household members performing activities, (c) spatial connectivity constraints on the vehicle, (d) spatial connectivity constraints on the household members, (e) vehicle and household member coupling constraints, and (f) capacity, budget and participation constraints.

2.1 Objective Function

The model generates the optimal solution given an objective of minimizing travel-related disutilities and assess the effect of using AVs on household members’ travel patterns. Thus, we consider the undesirable terms that may affect the travel decisions during the day and minimize the travel disutility caused by them in order to generate an acceptable travel pattern. Equation 1 shows the current objective function which consist of AVs’ travel costs, household members’ travel times and waiting times respectively. Other travel related decisions such as AVs parking cost, household members extent of the day, return home delay caused by trip chaining, etc can also be considered. In order to keep the objective function linear, a travel disutility function is estimated to be a linear combination of terms. (Chow and Recker, 2012; Kang and Recker, 2014; Regue, 2015; Recker et al., 2008).

\[
\text{Min} \ \ \beta_1 \sum X_{i,j}^v t_{ij} + \beta_2 \sum H_{i,j}^p t_{ij} + \beta_3 \sum W_j^p
\]

\(\beta_s\) indicate the proportional effect of the corresponding travel disutility term in the total objective function value. For example, the currently known time value of traveling, $10/h-$30/h, is expected to be significantly reduced by using AVs (Small et al., 2005). Since the “hands-free” travelers may make other productive use of the time while traveling, thus becoming more tolerant of travel times and distances. A sensitivity analysis of varying time values of travel can identify the trend of effects on VMT, number of trips, etc.
2.2 Computational Methods

The proposed model integrates two NP-hard routing and scheduling problems and this integration adds an enormous burden on the computation. In addition, the fact that travelers’ cost is comprised of time decisions as well as path decisions (i.e., \( Z = f(X, H) + f(T) \)), makes the problem extremely complicated (Kang and Recker, 2013) compared to the vast majority of VRPs where the objective function only depends on path decisions (i.e., \( Z = f(X) \)). Thus, commercial solvers such as CPLEX may solve only small size HAPP-AVs and heuristic or decomposition methods should be employed to solve the problem.

We present two computational methods to solve HAPP-AV. The basic idea of both methods is breaking down the entire problem into two parts. In the first step, we find activity assignments for every household members while we don’t consider any vehicle constraints in this step. In the second step, using the output assignments of the first step, we find the best AV routes to satisfy the household members’ travel needs.

The first method generates the optimal solution of the original problem. We generate all possible assignments for each household member in the activity assignment step, the first step. However, we also employ pruning rules to eliminate some infeasible activity-human combinations. In the second step, we solve the AV routing problem for all remained activity assignments. Using bounds of the AV routing problem, some of the problems can be fathomed. Finally, we choose the activity assignment and corresponding AV route with the minimum total objective function value. This method is much faster than using CPLEX directly to solve the entire problem, which we discuss in section 2.4; however in case of very big problems, we introduce the following heuristic.

The second method is a heuristic in which we only consider the best (most desired) activity assignment for each household member and find the best AV route for that assignment in the second step. This approach is based on the logic that people would perform their daily activities based on their own willingness other than vehicle constraints. However, this approach may result in an infeasible solution which we take care of it by generating near optimal alternative solutions in the first step; The solution may not even be close to the optimal solution of the original problem which has a combined human/vehicle objective function.

2.3 Illustrative example

As an example of the application of HAPP-AV we consider the case of a household with two household members and a autonomous vehicle. The household members have three activities, (1) work, which should be done by the first member, (2) a social activity, which should be done by the second member and (3) shopping which can be done by either of the household members. The information for the activities is given in table 1. The social activity location does not have a parking location but other activities have their own parking locations.

The HAPP-AV model suggests activity patterns and AV routes which are shown in figure 2 to minimize household disutility. Dashed lines represent waiting time and empty trips. Person 1 performs work and shopping and person 2 only performs the social activity. Both household members shared the AV in the beginning of the day. The AV drops off the first person at work at 9 and then person 2 at the social activity location at 9.5. However, the household member has to wait for half an hour because of the activity time window. Since there is no parking available at the social activity location, the AV goes back home to park. The AV leaves home again at 12.5 to pick up person 2 and brings...
Table 1: Activities information

<table>
<thead>
<tr>
<th>Activity</th>
<th>Duration</th>
<th>Arrival time window</th>
<th>Actual activity start time window</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work - P1</td>
<td>5</td>
<td>[8,9]</td>
<td>[9,9]</td>
</tr>
<tr>
<td>Social activity - P2</td>
<td>3</td>
<td>[9,11]</td>
<td>[10,11]</td>
</tr>
<tr>
<td>Shopping</td>
<td>1</td>
<td>[12,18]</td>
<td>[14,18]</td>
</tr>
</tbody>
</table>

him/her back home. In the mean time, person 2 finishes his work at 14 and waits to be picked up. The AV picks him up at 14.5 and drops him at the shopping location at 15 and parks there. Person 2 finishes shopping at 16 and the AV picks him up and arrives home at 16.5.

Figure 2: Household members and AV travel pattern

2.4 Case studies

We use the California Statewide Travel Survey for households that reside in Orange County and Los Angeles County to assess using one AV instead of their regular vehicles. A total of 86 households that have only two adult household members are selected. Originally, these households used two conventional vehicles to serve their travel needs. HAPP-AV is solved for all the instances by implementing the first solution approach and using CPLEX directly. The summary of the results is shown in table 2.

We had a 30 hour time limit for solving each instance by CPLEX. Only 55% of the instances were solved (3 of them were infeasible) by CPLEX in the time limit. However 88% of the instances were optimally solved by the first solution approach, 11% were infeasible and only 1% could not be solved.
This outcome is interesting since only 11% of the households require more than one AV to perform their daily activities that was served by two conventional vehicles previously. There is also a huge improvement in the runtime by using the new solution approach. The optimally solved instances take almost 2.5 hours on average to be solved by CPLEX, however it only take less than 10 minutes by using our solution approach.

<table>
<thead>
<tr>
<th></th>
<th>Number of all instances</th>
<th>Number of optimally solved</th>
<th>Average first step runtime</th>
<th>Average total runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPLEX</td>
<td>86</td>
<td>48</td>
<td>-</td>
<td>2.4 hr</td>
</tr>
<tr>
<td>Solution approach 1</td>
<td>86</td>
<td>75</td>
<td>0 sec</td>
<td>8.2 min</td>
</tr>
</tbody>
</table>
References Cited


Estimating Primary Demand of One-Way Vehicle Sharing Systems

Chiwei Yan and Chong Yang Goh

Operations Research Center, MIT

1 Introduction

One-way vehicle sharing systems, where a driver can pick up a vehicle and drop it off later at a station closest to his or her destination, are being widely adopted all over the world: according to (Larsen 2013), one-way bike-sharing systems have grown from 213 services operating in 14 countries with 73,500 total bikes in year 2008, to over 500 services with more than 500,000 bikes in year 2014; one-way car-sharing systems are also gaining popularity, as evident in the increasing number of car rental companies launching one-way services (Enterprise Car Share, Hertz on Demand, Zipcar, Car2go).

A primary challenge in operating such systems is to maintain both pick-up and drop-off availability at various service locations. This requires proper balancing of supply and demand, which is usually done by periodic relocation of vehicles from station to station. For example, in a bike-sharing system, a station contains a fixed number of bike docks, each of which is either occupied (available for pick-up) or empty (available for drop-off) at a given moment. When pick-up and drop-off demands are asymmetric, which commonly occurs during peak hours, available bikes in the system need to be reshuffled from time to time to ensure the appropriate number of bikes and empty docks are available at each station. Many works have been done in developing models and algorithms to optimize these rebalancing operations (O’Mahony and Shmoys 2015, Raviv et al. 2013, Shu et al. 2013). Some have been successfully implemented in real-world systems, such as Citibike in New York City (O’Mahony and Shmoys 2015).

Crucially, these rebalancing methods rely on having accurate estimates of primary (first-choice) demands for pick-ups and drop-offs at each station over time. But as acknowledged in some of these papers, this estimation problem is non-trivial. To illustrate why, consider an example in Figure 1 showing an area in Cambridge with three bike stations (named A, B, C) and a subway station (denoted with a circled T). If an arriving customer intends to pick up a bike at Station A but finds it empty, he or she may consider several alternatives: (1) go to either station B or C instead to pick up a bike, if available (2) take the subway and leave the system. We call this the demand substitution effect. An
implication is that if we simply aggregate the observed trip data, we may underestimate the first-choice demand at a station with frequent stockouts (Station A) and overestimate demands at its nearby stations (Station B and C).

Figure 1: An illustration of demand substitution in bike-sharing systems

Most existing works have focused heavily on the operational side of the rebalancing problem, while adopting relatively naïve methods for estimating the primary demands. For example, in O’Mahony and Shmoys (2015), the authors proposed to first filter the data by dropping time periods where a station is empty, and then estimate the demand by a simple aggregation procedure. As we will illustrate in Section 3, this can result in significant biases. On the other hand, by explicitly modeling the demand substitution behaviors, one may expect to reduce such biases and even recover insights on how substitutions occur.

The focus of this paper is to develop a rigorous estimation approach with explicit modeling of demand substitution behaviors in order to recover primary demand for each station, using observed trips and station availability data. Our contributions are as follows:

1. To the best of our knowledge, ours is the first proposed approach for estimating station-wise primary demand in a bike-sharing system, by taking into account of complex and realistic substitution behaviors.

2. Aside from station-wise primary demand, the proposed approach also learns the preference structures of commuters from data. The preference structure could provide insights into how frequently substitution occurs, including the fractions of commuters that may leave the system due to service inavailability.

2 Methodologies

Assume that there are $n$ stations in total, and commuters arrive according to a homogeneous Poisson process with rate $\lambda$. We model demand substitution
behaviors using a ranking-based choice model (Farias et al. 2013, van Ryzin and Vulcano 2014). Each arriving commuter is assumed to have a preference ranking over all stations, including an option to leave the system if the stations are not available (i.e., empty at pick-up or full at drop-off). For example, in Figure 2, the arrived customer has a preference ranking of “station 3, station 2, leave, ...”. This preference ranking completely determines the commuter’s behaviors as follows: if all stations are available, this commuter will pick a bike at station 3, which is the first choice. If station 3 is stocked out, the commuter will consider the second choice at station 2. However, if both stations 2 and 3 are stocked out, the commuter will leave the system, which is the third option. We denote the probabilities over all such possible rankings in a vector $x$, where $x_i$ is the probability that preference ranking $i$ will occur. Therefore, the dimension of $x$ is the total number of preference rankings considered. We assume that each arriving commuter has a random i.i.d. draw of preference ranking according to the multinomial distribution specified by $x$, and is independent from each other.

![Figure 2: An illustration of the ranking-based choice model](image)

The goal of the estimation task is to recover $x$ from data. We take a frequentist approach here, where we first construct the likelihood function from data based on the aforementioned probabilistic model, and then solve for $x$ using Maximum Likelihood Estimation (MLE). Due to space constraints, we omit the details of the derivation of our likelihood function. Instead, we want to point out that the key challenge of applying MLE to our problem is that the dimension of $x$ is too high. The number of all permutations of $n$ station along with the leave-the-system option is $(n + 1)!$. In a large bike-sharing system, $n$ is usually in the hundreds. Existing estimation methods in (Farias et al. 2013, van Ryzin and Vulcano 2014) do not scale well to this problem size.

To address this issue, we developed an implicit enumeration technique to restrict the set of feasible permutations, by only including “meaningful” preference rankings that satisfy certain practical constraints. This is done by exploiting two structural properties of bike-sharing systems: (1) commuters will usually consider only nearby stations for substitutions; (2) the fraction of stations stocked out at each time period is rather small in practice. With this approach, we are able to prune the set of feasible preference rankings down to several thousands in a city-scale bike-sharing system. This makes our approach suitable for implementation in real-world instances.
3 Preliminary Experiments

We did preliminary numerical experiments using data from a bike-sharing system in Boston (Hubway) consisting of 53 stations. We only look at pick-up demand for now. The drop-off demand can be estimated using almost the same approach. We developed a simulation platform to generate synthetic trip data, which is based on the actual bike-sharing network. We assume there is an underlying ground truth primary demand for each origin-destination station pair. The primary demand rate is calibrated using actual trip data from 5:00 PM to 7:00 PM for 5 weekdays. We assume two different substitutions in the simulator:

1. *Destination-independent substitution*: users arriving at its first-choice station with no bikes will leave the system with probability 0.3. With probability 0.7, it will search nearby stations in ascending order of the distance between the alternative station and the first-choice station, and this distance cannot exceed 1km.

2. *Destination-dependent substitution*: users arriving at its first-choice station with no bikes will leave the system with probability 0.3. With probability 0.7, it will search nearby stations in ascending order of the sum of the distance between the first-choice station and the alternative station, and the distance between the alternative station and its intended destination station. Again, the distance between the first-choice station and the alternative station cannot exceed 1km.

We benchmark against two baseline approaches:

1. *Sample Average*: average over all observed trip demand over all time periods.

2. *Filtered Sample Average*: average over all observed trip demand for only time periods where the station bike is non-empty. (O’Mahony and Shmoys 2015)

We use the simulator to generate 30 hours of synthetic trip data under two substitution behaviors. We then run our proposed estimation approach along with the two baselines to recover the station-wise primary pick-up demand. For the proposed approach, we test the estimation results with different maximum length of the preference ranking, which is defined as the number of stations before the leave-the-system option. We report the mean absolute error (MAE) of the estimation error for different approaches. The results are summarized in the following Figure 3. We can see that our method has a clear edge when substitution behaviors are complex (destination-dependent substitution). And in general, better performance is expected when we allow longer length of preference rankings.
4 On-going Work

We are currently working on experiments with real trip and station availability data. We are also developing a model selection procedure based on AIC/BIC criteria so that only the best subset of preference rankings will be considered in the model to avoid over-fitting and improve model generalization performance.

References


DIFFERENCES BETWEEN USERS AND NON-USERS OF CROWDSOURCED SHIPPING

Aymeric Punel¹, Alireza Ermagun², and Amanda Stathopoulos³

¹Ph.D. Student, Department of Civil and Environmental Engineering, Northwestern University
²Postdoctoral Fellow, Department of Civil and Environmental Engineering, Northwestern University
³Assistant Professor, Department of Civil and Environmental Engineering, Northwestern University

Corresponding Author Email: a-stathopoulos@northwestern.edu
1 INTRODUCTION

Logistics operators are currently facing the challenge of addressing evolving service expectations among consumers. Amidst growing volumes of goods delivery operations due to the diffusion of ICT and e-commerce, customers are also demanding transparency, reduced cost, and shorter shipment times. It is consequently essential for city logistics providers to adapt to the market evolution and satisfy the increasing demand related to e-commerce and same-day delivery.

There is increasing interest in freight delivery innovation related to the use of crowdsourced vehicles and drivers. Examples include DHL’s MyWay experiment, a last-mile delivery scheme where e-purchased items were picked up and delivered by neighbors. Moreover, several start-up companies are specializing in shipping via the crowd, such as Zipment, Roadie, and Nimber. Crowdshipping is built on the notion that non-professionals (like everyday commuters) transport packages along their planned routes. A delivery system based on crowdshipping, given its reliance on existing resources, has the potential to reduce the environmental impact by optimizing deliveries and reducing the total number of vehicles on the road. Ranging from urban deliveries to international shipping, crowdshipping can be applied to many different business models, each one with its own characteristics, and user motivations. Nevertheless, because of its short existence, it raises many questions concerning its operational capacities, and the extent to which it can be beneficial to the society remains unanswered.

While crowdshipping alleviates congestion by reducing the number of vehicles on the road, it may also produce more mileage as commuters make detours to pick-up and deliver packages. In addition, it may encourage people to ship goods that would not have been sent otherwise, creating additional flows. Furthermore, some companies also claim that they will be able to build a strong community among their users, based on sharing and collaboration, whereas users might not pay attention to this argument and are only looking for a cheap delivery service. These contradictions make the study of crowdshipping a challenging problem. This paper aims to analyze crowdshipping user behavior and motivations to begin building a knowledge framework around the public view and use of an innovative goods delivery systems. Empirical analysis of the individual and contextual factors that promote crowdshipping acceptance is essential to improving operations and underlying business models of crowdshipping as well as designing policies to decrease the negative impacts related to freight movement. The objective of this study is to understand whether and to what extent the attitudes and expectations of crowdshipping users differ from non-users. Not only does it enable us to pave the crowdshipping road to find its niche, but it also adds to the fundamental insights on crowdshipping systems. The rest of the abstract is as follows. First, we provide a description of the data used in this study. Second, we discuss the methodology used for the analysis followed by the initial results of the models.

2 DATA

This study is based on an online survey conducted among 800 people in four US states, namely, California, Florida, Georgia, and Illinois in June 2016. The questionnaire was divided into five sections:

1. The first section covers questions to determine respondents’ previous experience with the parcel and crowdshipping industries. This section includes questions such as “Before taking this survey, were you familiar with the concept of crowdshipping?” and “On average, how many packages are you sending through a crowdsourced delivery service each year?”

2. The second section presents a series of stated choice scenarios setting respondents in the situation where they have to send a package through a crowdshipping platform.
3. The third section asks respondents about their opinion on the parcel industry and their preferences regarding this service. This section encompasses close ended questions such as “I would like to try a crowdshipping service because it allows me to save money,” “I would like to try a crowdshipping service because it is more convenient than traditional delivery,” and “I have reservations about trying crowdshipping service because I wouldn't trust the drivers.” Respondents were asked to rank these questions from “Strongly Disagree” to “Strongly Agree” on 7-point Likert scales.

4. The fourth section presents statements related to respondents' attitudes towards using crowdshipping, including “I believe crowdshipping is a platform where its users, both senders and drivers, can help each other” and “Crowdshipping provides advantages for both drivers and senders.” Similar to the third section, the questions were designed as 7-point Likert scales.

5. The fifth section asks respondents about some demographic and identity information including age, gender, and level of income.

Respondents were selected from both crowdshipping users and non-users, as the survey aimed to understand the potential shippers from the general public and their evaluation of crowdsourced delivery options. In total, 587 people completed the survey with a return rate of 73.4%. Following the data cleaning process, which includes removing bad and incomplete observations along with excluding outliers, 533 observations are retained for further analysis.

3 METHOD AND MODEL

To analyze the differences between crowdshipping users and non-users, we employ two distinct methodologies: (1) A Wilcoxon Rank-Sum test, and (2) A Probit model. The former aims to analyze whether and to what extent the attitudes and expectations of the crowdshipping users differ from non-users. The latter tries to understand how the crowdshipping users are distinguished from non-users. The following subsections elaborate the methods in detail.

3.1 Wilcoxon Rank-Sum Test

Wilcoxon Rank-Sum test is employed to detect whether there are differences in behavior between the crowdshipping users and non-users. It is applied to the series of positive and negative statements about crowdshipping along with the attitude statements extracted from the fourth section of the survey. Wilcoxon Rank-Sum test is a nonparametric alternative test to the unpaired t-test. This test is typically used when the normality assumption of sample values is violated. To choose the appropriate method, we tested the normality assumption with Shapiro-Wilk Normality test. We then selected Wilcoxon Rank-Sum test, followed by the rejection of the normality assumption.

The first results of the Wilcoxon Rank-Sum test reveal differences between crowdshipping users and non-users. All respondents agree that crowdshipping is an easy system which brings advantages and is more efficient than traditional shipping. However, people who have already used crowdshipping are more aware of the potential benefits of crowdshipping related to sustainability values than non-users.

3.2 Probit Model

To distinguish the crowdshipping users from non-users, we employ a Probit model. We test a wide variety of variables, encompassing socioeconomic, demographic, attitudes and preferences, and built environment variables. Pertaining to the socioeconomic and demographic variables, gender and age
were found significant in the model. The results indicate that women are less likely to use crowdshipping. As expected, the probability of using the crowdshipping service among the people older than 45 years is less than people below this age. Pertaining to the attitudes and preferences, the results show that users are interested in crowdshipping systems as it is environmental friendly. We also reveal that non-users strongly agree that the untrustworthiness of drivers is one of the main reservations related to crowdshipping. Interestingly, we found users agree that the crowdshipping is complicated to use.

**Keywords:** Freight transportation; Crowdsourced shipping; Mobility-as-a-service; Behavioral survey
Travel demand estimation using heterogeneous data pieces: addressing stochasticity and observability issues

Yudi Yang\textsuperscript{a}, Yueyue Fan\textsuperscript{a,\ast}, Roger Wets\textsuperscript{b}

\textsuperscript{a}Department of Civil and Environmental Engineering
University of California Davis, CA 95616
\textsuperscript{b}Department of Mathematics
University of California Davis, CA 95616

\textbf{Keywords:} travel demand estimation, observability, identifiability, heterogeneous data

In the past decades, most of static O-D demand estimation studies focused on generating a single vector of estimate for all pairs that is either considered as constant demand throughout the entire estimation horizon or the mean of stochastic demand across the days. However, in order to simulate the most realistic road network, a more comprehensive understanding of stochastic travel demand is desired for a large variety of reliability based transportation analyses. For example, stochastic traffic network models, such as stochastic network design \cite{Faturechi2014, Lai2013, Santoso2005, Rezaee2015} and network reliability models \cite{Clark2005, Shao2006, Lam2008, Chen2010} all require knowledge of the uncertainty of travel demand as model input. The list of publications that needs stochastic demand as input could go much longer. However, the scope of literature devoted to the understanding of stochastic travel demand is rather limited. In this paper, we address two major challenges involved in stochastic demand estimation: (1) how to effectively integrate heterogeneous data and knowledge pieces over a network structure; (2) how to overcome the observability/identifiability issue (i.e. to ensure unique solution) that is often inherited in network demand estimation problems.

Substantial efforts were made by Hazelton and coauthors in improving statistical approaches designed for O-D estimation, which includes differentiating problems of reconstructing, estimating, and predicting an O-D matrix \cite{Hazelton2001}, developing Bayesian approaches for
inferring time-varying O-D matrix \cite{Hazelton2008}, and incorporating day-to-day dynamics brought by travelers learning \cite{ParryHazelton2013}. Most studies consider only hard data collected from fields, transportation science domain knowledge such as traveler’s route choice behavior was typically not incorporated in the estimation problems, except a recent study by \cite{Shao2014}, which proposed a bi-level program to estimate the mean and variance of multivariate normally distributed travel demand based on link counts. In their bi-level model, the upper-level minimizes the estimation error and lower-level imposes reliability based traffic user equilibrium.

Another important line of efforts in the literature of travel demand estimation focuses on the issue of solution uniqueness. The term observability is typically used for the determinacy of short-term travel demand from a single observation set, while the term of identifiability denotes the uniqueness of parameters that govern the probability distribution of travel demand using all input data. In brief, observability is associated with reconstruction problem and identifiability is with estimation problem. Although several studies have investigated the observability of travel demand based on nodal conservation and/or link proportion matrix \cite{Castillo2008} and \cite{Viti2014}, observability study on travel demand with the incorporation of traffic equilibrium is still lacking. In the context of stochastic demand estimation, additional distributional assumption was often assumed. For example by taking advantage of the unique feature of a Possession model, that is, mean being equal to variance, \cite{Hazelton2003} employed the second order property to address identifiability issue.

In this presentation, we will introduce a new stochastic demand estimation framework that enables more effective utilization of data and domain knowledge to mitigate identifiability issue. We will provide both theoretical analyses and numerical results to demonstrate the effectiveness of our approach. In addition, this new framework does not rely on any special probability or behavior assumptions, thus providing a flexible platform to incorporate a large variety of data/information types and user behaviors, which is especially valuable considering the potential technological and behavior transformation faced by our transportation system.

Next, we introduce the stochastic travel demand estimation problem. The probability distribution of the multivariate random variable for O-D demand \( \mathbf{X} = [X_i] \in \mathbb{R}^n \) is described by the parameters \( \theta \in \Theta \), where \( \Theta \) denotes the parameter space. Let \( \mathbf{Y} = [Y_j] \in \mathbb{R}^m \) and \( \mathbf{Z} = [Z_a] \in \mathbb{R}^{|A|} \) be random path flow and link flow respectively, resulted by demand \( \mathbf{X} \).

With the advancement of information and communication technologies, more and more traffic data will become available. In this paper, we consider two types of sensor data: link counts
and path counts. These sensor data are denoted by $W \in \mathbb{R}^S$, where $S$ is the number of data elements. For the entire estimation horizon, repeated observations from sensors form multiple sets of sensor data, $W^{(k)}, k = 1, ..., K$, where $K$ is the total number of observation periods. Let $V^{(k)}$ denote the true flow happened in observation period $k$ that corresponds to each element in $W^{(k)}$. The error of sensor data $\epsilon^{(k)} = W^{(k)} - V^{(k)}$ is considered to be relatively small and independent. There are several challenges associated with sensor data. Firstly, sensor data is often indirect, and the mapping between directly measurable parameters (link and path flows) and the parameters to be estimated (travel demand) may be complicated especially in a congested network. Secondly, sensor data may be incomplete, i.e., sensors may not cover all links and paths in a network due to practical and budgetary constraints. Lastly, sensor data alone may be insufficient to uniquely determine the travel demand.

In addition to sensor data, we also allow incorporation of target data $T$, which can be from a mix of household survey and road interview from current estimation horizon as well as historical estimates. We shall be cautious while using target data, because this data may be obsolete and unreliable, thus introducing bias to the estimates. Our modeling philosophy is that in the situation where target data has to be incorporated for observability reason, we shall maximize the information gain from sensor data and keep the dependence of estimates on unreliable target data minimal. This is why we need to construct an estimation framework that preserves the most information of sensor data, including not only the traffic count data itself, but also the physical and behavioral relationships among the counts reflected by transportation domain knowledge.

If demand variables are fully observable in all observation periods, a straightforward estimation method is to first reconstruct $X^{(k)}$ using $W^{(k)}$ for each observation period $k$ separately and then to estimate the parameter of $X$ based on appropriate probability model. Alternatively, some studies choose to first compress observed data to some sample statistics, and then compute the first order (mean) and sometimes second order (covariance matrix) of the demand based on simple (typically linear) relation between observed flows and the estimated travel demand.

The proposed framework integrates estimation and reconstruction for each period. On one hand, each reconstruction subproblem supplies $x^{(k)}$ as inputs to the estimator. On the other hand, estimation provides $E(X|\hat{\theta})$ to each reconstruction subproblem to ensure observability. In that, domain knowledge for all time periods is shared between the subproblems through estimation. Target data does not affect reconstruction directly as it happens in statistics-based problem. In that, we take advantage of good quality data and reduce the impact of poor quality.
The general formulation of stochastic demand estimation (SDE) problem is expressed as

\[
\text{SDE} : \min_{\theta \in \mathcal{Q}} h(\theta; t) + \sum_{k=1}^{K} s(x^{(k)}; w^{(k)}) + \sum_{k=1}^{K} r(\theta, x^{(k)})
\]

\[
s.t. \quad v^{(k)} = D y^{(k)}, z^{(k)} = \Delta y^{(k)}, x^{(k)} = \Gamma y^{(k)}, z^{(k)} = g(x^{(k)}), y^{(k)} \geq 0, \forall k = 1, \ldots, K
\]

where function \(h(\theta; t)\) measures the distance between \(E(X|\theta)\) and \(T\), \(s(x; w^{(k)})\) penalizes the difference between \(v\) and \(w^{(k)}\), and function \(r(\theta, x^{(k)})\) penalizes the difference between the estimated and reconstructed demand. The constraints define the physical connectivity of the network, and the mapping \(g(\cdot)\) captures user-network interactions. For example, \(g(\cdot)\) could represent a traffic user equilibrium model. From Figure 1, one can clearly see that SDE framework differs from the other two methods in the feedback of \(E(X|\theta)\) towards reconstruction and the use of \(T\) on estimation.

In the remaining part of the paper, we first analyze statistical convergence of the proposed estimation framework. We then provide proof of improved identifiability property of this new method. Finally, we provide numerical examples using Anaheim network (in 1992), which consists of 38 zones, 416 nodes, and 914 links. Both analytical and numerical results demonstrate improvement in the estimation quality of our method.

References


Supply Chain Logistics & Methods
SB4: City Logistics and Inventory Control
Saturday 11:15 – 12:45 PM
Session Chair: Teodor Gabriel Crainic

11:15  Value Function Approximation-based Dynamic Look-ahead Policies for Stochastic-Dynamic Inventory Routing in Bike Sharing Systems
Jan Brinkmann, Marlin Ulmer, Dirk Mattfeld*
Technische Universität Braunschweig

11:45  Synchronizing City Logistics with Sliding Time Windows
Saijun Shao*, Gangyan Xu, Ming Li, George Q. Huang
The University of Hong Kong

12:15  Multi-Modal Scheduled Service Network Design for Two-Tier City Logistics System with Resource Management
1Pirmin Fontaine*, 1Teodor Gabriel Crainic, 1Ola Jabali, 1Walter Rei
1ESG, UQAM & CIRRELT, 2DEIB, Politecnico di Milano
Value Function Approximation-based Dynamic Look-ahead Policies for Stochastic-Dynamic Inventory Routing in Bike Sharing Systems

Jan Brinkmann · Marlin W. Ulmer · Dirk C. Mattfeld
Technische Universität Braunschweig, Germany

1 Introduction

We consider a station-based bike sharing system (BSS), where users can rent and return bikes spontaneously. Stations are distributed in the city. BSS are particularly used by commuters and/or as a complement to public transportation. We typically observe one-way trips and rental and return requests do not occur at the same station. Further, rental and return requests are uncertain and subject to spatial-temporal patterns (Borgnat et al., 2011; Vogel et al., 2011; O'Brien et al., 2014). The success of BSS strongly depends on two factors: stations’ density within the city and a reliable availability of bikes and free bike racks at any time and any station (Gauthier et al., 2013). Therefore, service providers dynamically dispatch transport vehicles to relocate bikes between stations. The challenge is to balance the number of bikes and free bike racks at every station to satisfy as many requests as possible.

We model the rebalancing problem as an inventory routing problem (IRP). In IRP, customers consume a certain commodity provided by vehicles. The goal is to prevent the customers’ inventories from becoming void (Coelho et al., 2014b). In BSS, two dependent resources, namely bikes and free bike racks, need to be balanced. The number of bikes within the BSS and bike racks at every station are fixed. Further, the availability of bikes and free bike racks at every station are interdependent. Therefore, standard IRP formulations are not applicable. Another challenge for the considered problem is that rental and return requests are uncertain and subject to a spatio-temporal stochastic pattern. This results in a stochastic IRP. Finally, due to the high variability of requests, sudden imbalances may occur on short notice. Thus, dynamic decision making is applied to take advantage from revealed information and to react to sudden imbalances. As a result, we consider a stochastic-dynamic IRP with interdependent inventories (SDIRP).

For the SDIRP, a set of vehicles and a set of capacitated stations with initial fill levels are given. Over the day, customers stochastically request rentals or returns of bikes at stations. To provide a sufficient number of free bikes and bike racks on each station, vehicles are dynamically dispatched to transport bikes between stations. Decisions are made about the inventory of the station a vehicle is currently located and about the next station the vehicle visits. A request is called failed if a bike (to rent) or bike rack (to return a bike) is not provided at the station and the time the customer requests. In case of a failure, the customer approaches a neighboring station for a new request. The objective is to minimize the expected number of failed rental and return requests.

We formulate the SDIRP as a Markov decision process (MDP, Puterman, 2014). The challenge for the SDIRP is that the costs of the MDP resulting from failed rentals and returns are not known when decisions are taken. For an evaluation of the decision’s outcome, anticipation of future customer requests is necessary. Thus, we derive anticipatory policies by means of approximate dynamic programming (ADP, Powell 2011). We introduce limited look-ahead
policies (LA) simulating future customer requests over a limited horizon to evaluate feasible inventory and routing decisions. These simulations are conducted in real-time (online). The anticipation’s accuracy strongly depends on the length of the simulation horizon and on the request pattern. If the horizon is too short, only a few request can be observed and thus no anticipation is possible. If the horizon is too long, the anticipation becomes distorted due the spatio-temporal pattern of requests. This distortion may vary over the day since it depends on several factors, particularly the request frequency and the trip purpose of requests. Commuter behavior may be easier to anticipate than more random leisure requests. Thus, the suitable length of the simulation horizon may vary over the course of the day. To this end, we develop dynamic LAs (DLA) with time-dependent simulation horizons. To determine time-dependent simulation horizons, we parametrize the DLA by means of value function approximation (VFA, Powell 2011). First, we aggregate states to periods. Then, offline simulations are carried out to approximate the expected numbers of future failed requests, given a combination of a period and a specific simulation horizon. Eventually, we achieve a policy with individual horizon per period. In a comprehensive computational study based on real-world data of the BSS in Minneapolis (Minnesota, USA, Nice Ride MN 2016), we show that VFA-based DLAs significantly outperform LAs with static horizons as well as benchmark policies from the literature by Coelho et al. (2014a) and Brinkmann et al. (2015). The high solution quality is enabled by the DLA’s adaption to the request pattern.

Our contributions are as follows. For the bike-sharing application, we present a comprehensive MDP-model for the stochastic-dynamic IRP for station-based BSS. We develop anticipatory methods autonomously adapting to the instance’s characteristics and significantly reducing the number of failed rental and return requests compared to benchmark policies from the literature. Methodologically, we present a new and general solution method, by using VFA to assign a policy to a subset of states. This can be seen as non-parametric policy search (compare Powell 2011, pp. 249ff. for an overview of parametric policy search). Given a set of candidate policies, this method allows a policy selection based on state-parameters. This policy selection is guided by VFA. Non-parametric policy search may provide value particularly for stochastic-dynamic decision problems of high complexity, often experienced in the fields of transportation and routing.

In the remainder of this abstract, we briefly present the MDP for the SDIRP in Section 2. In Section 3, we further describe the solution method in detail. The setting and the results of the computational studies are presented in Section 4.

2 Stochastic-Dynamic Inventory Routing for Bike Sharing Systems

In the following, we present the MDP-model for the SDIRP. Decisions are made over a sequence of decision points \( K = (0, \ldots, k_{\text{max}}) \). A decision point \( k \) occurs when a vehicles arrives at a station. The associated decision state \( s_k \in S \) involves information on the point in time, stations’ fill levels, vehicles’ positions and loads, and travel times. Given a decision state \( s_k \), decisions \( x \in X(s_k) \) are made about the number of bikes to pick up or deliver as well as about the next station to visit for the vehicle. When a decision is made, a stochastic realization \( \omega_k \in \Omega \) is revealed. The stochastic realization provides a set of user requests for every station. Satisfied requests alter the fill levels at the stations. Unsatisfied request lead to costs \( c(s_k, x, \omega_k) \in \mathbb{N}_0 \). The subsequent decision state \( s_{k+1} = (s_k, x, \omega_k) \) occurs when the next vehicle arrives at a station.

A policy \( \pi : S \to X \) maps a state \( s_k \) to a feasible decision \( x \in X(s_k) \subseteq X \). The objective is to identify an optimal policy \( \pi^* \in \Pi \) minimizing the expected number of failed requests, as depicted in Equation (1).
\[ \pi^* = \arg \min_{\pi \in \Pi} \mathbb{E} \left[ \sum_{k=0}^{k_{\text{max}}} c(s_k, \pi(s_k), \omega_k) \big| s_0 \right]. \]  

(1)

3 Dynamic Look-ahead Policies

To determine \( \pi^* \), the Markov decision tree needs to be solved exactly. For the SDIRP this is impossible due to the three curses of dimensionality (Powell, 2011). First, the state space is huge due to the large number of attributes describing a state, particularly the fill level at each station. Second, we experience a large decision space due to the combination of inventory and routing decisions. Third, the transition space is large due to the large number of potential request combinations. To this end, we aim on approximating Equation (1). We draw on two steps. First, we reduce the set of policies to DLAs. Second, we apply value function approximation to determine the best DLA.

As policies, we only consider dynamic look-aheads \( \pi_h \in \Pi' \subset \Pi \). Given a state \( s_k \), the DLA simulates a limited time \( \delta \) into the future and calculates the observed failed rentals and returns per station. Based on these results, the DLA determines the inventory decision at the current station and the next station to visit. The next station is the station where the highest number of simulated requests fail. To vary the simulation horizons, we first determine fixed periods \( \rho_i \in P \) of the time horizon (in our case hours). For every period, a simulation horizon \( \delta \in \Delta \subset \mathbb{N}_0 \) is selected. Therefore, a policy \( \pi_h \in \Pi' \) can be described as a sequence of simulation horizons \( h = (\delta_0, \delta_1, \ldots, \delta_{\text{max}}) \in H \) with horizon \( \delta_i \) for the LA applied in period \( \rho_i \). The overall set of candidate policies \( \Pi' \) is therefore same size as the Cartesian products of all horizon combinations \( h \in H \).

To determine a suitable sequence of simulation horizons, we draw on VFA. To this end, we approximate the expected costs to go \( \tilde{\nu}(\rho, \delta) \in \mathbb{R}^+_0, \forall \rho \in P, \delta \in \Delta \) by means of forward simulation. Then, we apply the approximate Bellman Equation (2).

\[ \delta_i = \arg \min_{\delta \in \Delta} \tilde{\nu}(\rho_i, \delta) \]  

(2)

Given a period \( \rho_i \), we solve Equation (2) to select a simulation horizon \( \delta \).

For the sake of space in this abstract, we omit further details of the VFA-procedure. We point out that due to the small number of 24 periods, an exploration procedure is obligatory. Thus, we draw on Boltzmann exploration (Powell 2011, pp. 467f.).

4 Computational Studies

To evaluate our policies, we draw on real-world data offered by Minneapolis’ (Minnesota, USA) bike sharing system ”Nice Ride MN” (Nice Ride MN, 2016). After data preprocessing, the data set comprises 161 stations, 88 working days in the summer month of 2015, and 197,726 trips. On average 2,246.89 trips occur per day. By randomly drawing trips, we can sample requests and preserve the spatio-temporal pattern. Every instance comprises one working day. A day is subdivided into 24 periods, each of 60 minutes. Figure 1 depicts the temporal distribution of trips in the course of the day. In periods 8 and 17, we observe two peaks due to commuters. An other peak appears at noon during lunchtime.

We consider the VFA-parametrized DLAs with \( \delta \in \{0, \ldots, 6\} = \Delta \). Horizon \( \delta = 0 \) results in a myopic policy we reinforce with static safety buffers of bikes and free bike racks as proposed in Coelho et al. (2014a) and Brinkmann et al. (2015). As benchmark policies, we compare DLA with LAs of static simulation horizons and the aforementioned myopic base.
policy. In our computational setting, the policies dispatch one vehicle. For an evaluation of the approaches, we simulate 1,000 non-concatenated working days.

The best static LA’s (SLA) simulation horizon comprises $\delta = 3$ periods. DLA reduces failed requests compared to the myopic policies by 41.2%. This shows how anticipation allows significant reductions of failed requests and increases customer satisfaction. Compared to the static LA, the DLA reduces failed requests by 10.8%. This indicates that an individual simulation horizon per period is highly beneficial. We show the horizons in Figure 2. On the left side, the simulation horizon for the static LA is shown. The y-axis depicts the current period. The x-axis depicts the period the simulation horizon ends. Since the horizon is static, we observe a constant horizon of three hours. The DLA shows a different behavior. We observe two significant periods until the DLA simulates, periods 8 and 17. These periods reflect the commuters peak hours as depicted in Figure 1. The VFA learns to simulate up to but not further than these peak periods. A simulation up to these periods is necessary to anticipate the commuters requests. In these periods, the number of requests is high and the transitions highly stochastic. An anticipation further than these periods leads therefore to inaccuracy and to inferior decisions. The VFA-based DLA is able to capture these phenomena adapting the horizons to the request pattern and makes decisions on accurate anticipation.
References


Synchronizing city logistics with sliding time windows

Saijun Shao, Gangyan Xu, Ming Li, G. Q. Huang*
HKU-ZIRI Lab for Physical Internet,
Department of Industrial and Manufacturing Systems Engineering,
The University of Hong Kong, Hong Kong

**Background:** Recent years have seen a roaring development of the E-commerce market. As a consequence, customers are now raising requirements for more satisfactory city logistics while at cheaper prices. How to balance the service level and costs becomes the key for enterprises. This research tackles this challenge by introducing a novel type of time constraint, the *sliding time window* (STW), to achieve satisfactory logistics services while maintain reasonable costs.

For over decades, researchers use time windows (TW) to temporally restrict transportation tasks. A tight TW would benefit customers with shorter waiting time and lower uncertainty, thus contribute to better customer satisfactions. On the other side, nevertheless, tight TWs would generally result in extremely high logistics costs, or even infeasibility. This conflict is particularly common in E-commerce markets, where customers wish to receive deliveries simultaneously after purchasing products from multiple vendors. It is rather costly (or even impossible) to achieve such synchronization by simply imposing a tight TW on all deliveries to the same customer.

The major innovation of STW is to relax TWs by allowing them to *slide* back and forth on time axis, while at the same time keep a tight size of TWs. This approach is inspired by our industrial collaborator, China’s leading furniture E-commerce platform operator. According to their practical experiences, customers strongly desire to receive all products within a tight TW to avoid a long-time waiting; instead they are less concerned with whether the deliveries are to be made one day earlier or later. The essence of STW is to ease the congestion of transportation resources usage and hence reduce costs.

**The VRP-STW model:** The VRP-STW model is established base on VRP-TW models, and replace the traditional time window constraints with synchronization constraints. For each customer, all deliveries need to be made within a time window, of which is size is not allowed to exceed the maximum width of synchronization time window. The actual synchronization level of a customer is calculated as the time duration between his or her first and last deliveries. The objective is to optimize the overall (or equivalently, the average) synchronization levels of all customers.

**Solving methodology:** The sliding nature of STWs bring about great challenges for solving VRP-
STW. To efficiently generate acceptable solutions, this research has adopted the concept of Divide and Conquer (DC), which attempts to gradually fix STWs by decomposing the problem into smaller-sized sub-problems which are transformed to classical VRPs, and then tackle them with the Tabu Search (TS) algorithm.

To divide customers into clusters, the algorithm of K-mean is adopted. It considers both the locations proximity and the order similarity when classifying these customers. A feasible cluster will be generated and saved when 1) the total demands of the cluster do not exceed the total capacity of the fleet; 2) the cluster does not contain any pair of customers between which the travel time exceeds the size of the synchronization time window; and 3) a feasible solution is found by the TS module, satisfying all the constraints.

For each cluster, we propose to use a TS-based algorithm to solve the generated standard VRP-TW. The initial solution is derived with nearest neighbor algorithm. In order to obtain neighbor solutions, we consider three types of operators, namely intra-route reallocation, inter-route exchange and 2-opt exchange. We use aspiration criteria accept certain inferior neighbors so as to escape local optima. The TS processes will terminate after a pre-set number of iterations has elapsed since start or last improved.

Conclusions:

(1) The effectiveness of VRP-STW: By comparing the results against traditional VRP-TW models, we found that the VRP-STW model could successfully serve several times more orders given the same number of vehicles. The average order fulfillment cost is reduced by more than 30% in most instances, and up to 70%. The service levels are also improved in most cases solved, which is particularly remarkable when the number of customers are relatively large.

(2) The impact of order similarity weight in clustering: We carry out a set of experiments to study how the weight allocation in clustering would affect transportation costs. The results show the optimal weight of order similarity increases along with the number of vendors. When there are relatively few vendors, the customers should be clustered simply according to their locations; however when more vendors enter the E-commerce platform, the order similarity becomes more vital in clustering processes so as to save the average delivery cost per order. This insight is critical for managers when their business scope expands and more vendors are attracted to join the platform, the weight of order similarity shall also be adjusted accordingly.
(3) **The impact of maximum synchronization time window (Max-|S|):** We demonstrated that a larger Max-|S| could always reduce the average transportation cost, and this reduction is more prominent given a larger number of vendors (up to around 30%). Another interesting finding is that the actual service level is always much better than the promised service level. For example, when Max-|S| is 0.5 unit of time, the actual average |S| is smaller than 0.1 unit of time. Based on our observation, only in a few extreme cases the actual |S| of the customer would be close to the Max-[STW], for example those customers who purchase products from a large number of different vendors.

(4) **The impact of maximum service time (Max-T):** We have revealed that the change of Max-T does not notably influence the performance when the number of orders is relatively small (e.g. 100 orders), which is probably because orders can be served duly and will not be accumulated. However when the number of orders is relatively large (e.g. 500 orders), the average order fulfillment cost will be reduced when Max-T increases. It is noteworthy that even a slight expansion of Max-T, say from 1 day to 2 days, can realize significant reduction in cost (21%) and increase in percentage of orders to be served (22%). It seems surprising that when even more orders arrive (e.g. 750 orders), Max-T = 1 results in the minimum cost. However, this is essentially due to the fact that a large percentage (34%) of orders being discarded; by changing Max-T to 2 and 3 days, the serving rate increases dramatically to 86% and 89%, respectively. Again, we could see the trends of average cost and serving rate coincide with our previous observations.
Multi-modal Scheduled Service Network Design for Two-Tier City Logistics System with Resource Management

Pirmin Fontaine¹, Teodor Gabriel Crainic¹, Ola Jabali², and Walter Rei¹

¹School of Management, Université du Québec à Montréal, and CIRRELT, Montréal, Canada
²Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy

1 Introduction

The transportation of goods in urban areas is a complex activity. This complexity is amplified by the increasing urbanization. In 2014, 54% of the world’s population was living in urban areas. The United Nations are expecting an increase up to 66% until 2050 and the OECD 85% until 2100. This results not only in larger demands but also attributes to increasing the intricacy of distributions networks. Increasing e-commerce further strains distribution systems. In Europe, B2C commerce increased in the last years by more than 13% each year and in 2016 an increase of 12% is expected as well. To improve the distribution, new organizations and business models are developed under the topic of city logistics (CL). Integrating goods distribution with existing public transport infrastructure is one way to improve quality of living in cities and reduce congestion. By consolidating flows in and out of the city, CL can improve the utilization of the means of transport and reduce driving distances. This also leads to a more sustainable transportation network.

Due to its increasing importance, CL found more attention during the last years. Bektas et al., (2015) and Savelsbergh and Van Woensel, (2016) recently reviewed the CL literature and pointed out the potential for future research. In two-tier systems, the literature mostly considers a vehicle routing problem (VRP) in the second stage. The first stage is either modeled as VRP as well, resulting in the two-echelon VRP (Perboli et al., 2011), or as service network design (SND) problem (Crainic et al., 2009). However, especially the literature on planning models is still scarce. We build on the formulation of Crainic et al., (2009), where services are selected and operated over a planning horizon of several months. We include different transportation modes and outbound demand in our model. Although these are two key elements of CL, they are not considered in existing models yet. Moreover, we consider resources in our model and identify which resources can improve the flexibility of the system and which resources are the main cost drivers. Further, we develop an efficient solution method since so far none for SND in CL exists.

¹pirmin.fontaine@cirrelt.ca
The contributions of this work are as follows: (1) We define a two-tier city logistics model for in- and outbound logistics which considers different transportation modes and resource management, (2) we propose a solution method for solving the problem, and (3) we identify the cost and flexibility effects of the different resources in such a system.

2 Problem definition and mathematical model

The 2T-CL system consists of two layers with the goal to deliver goods from origin to destination (in the classical setting from outside the city to customers in the city). All arriving goods have to be shipped to consolidation centers at the border of the city. We will refer to them as external zones. After consolidation and sorting in the external zones, the goods are shipped via urban vehicles to so called satellites from where the final distribution to the customers is done by city freighters. The outbound demand uses the same facilities in reverse direction: City freighters pick up the goods at the customer to transport them to a satellite. From there, urban vehicles transport the goods to the external zones.

In this model, we are optimizing a cyclic schedule for the first-tier services and the corresponding resources. The resulting plan is built for the whole schedule length, which is, depending on the problem, half a day or even shorter. This plan is then used for several month. Only the assignment of goods to the services and the second stage have to be optimized when the actual demand is realized.

The considered schedule length is divided into \( t = 1, \ldots, T \) periods, where each period represents a small time interval. As in (Crainic et al., 2009), the period length is defined such that (1) at most one departure of a service from its external zone may take place during a period, and, (2) the unloading times for urban vehicles are integer multiples of the period length.

The set of external zones is given by \( \mathcal{E} = \{ e \} \) and the set of satellites by \( \mathcal{Z} = \{ z \} \). The external zones are divided into external zones for the different means of transport \( \mathcal{M} = \{ m \} \) (trucks, trams and subways) \( \mathcal{E}^m \) (analogously for satellites \( \mathcal{Z}^m \)). If different modes are available in an external zone \( e \in \mathcal{E} \), \( e \) is part of all relevant subsets. The same hold for satellites.

Let \( \mathcal{T} = \{ \tau \} \) be the set of urban-vehicle types and \( \mathcal{T}(m) \) the subset of each transportation mode. The corresponding capacity of each urban-vehicle type is given by \( u_{\tau} \) and the fleet size at each external zone \( e \) by \( n_{e\tau} \). For trucks we assume that there is only one compartment, for trams and subways the number of same sized compartments is equal to the number of doors of all cars. The compartment capacity is given by \( u_{\tau}^c \) and the number of compartments is \( n_{e\tau}^c \) (such that \( u_{\tau}^c n_{e\tau}^c = u_{\tau} \forall \tau \in \mathcal{T} \)).

Each satellite has a capacity of urban vehicles \( u_{\tau}^{zt} \), for each period. Moreover, for the satellites with different means of transport \( m \in \mathcal{M} \), a capacity \( u_{mzt} \) for each is assumed. For trucks this is the actual number of trucks, while for trams or subways the number of cars is the limitation. The number of goods which can be unloaded or uploaded is typically reflected by the number of city freighters which can stop at a satellite. We assume a general capacity limit \( u_{\tau}^{zt} \) of satellite \( z \) in period \( t \).
The goal of the model is to satisfy the demand, given by set $D = \{d\}$, of a set of customers $C = \{c\}$. Further, to distinguish inbound and outbound logistics, the set of demands is divided into the two disjunctive sets $D^I$ for $c2c$ and $D^O$ for $c2e$ demand. Each demand is specified by volume $\text{vol}(d)$ and the customer location. A fixed cost of $k^E(d, e)$ is applied for assigning demand $d \in D$ to external zone $e \in E$. These costs avoid the exclusive assignment of products to external zones, and can be seen as costs for using a tram in the surrounding of the city or further transportation costs for the carrier. The costs for the best located external zone from the carrier perspective might be even zero. In the first-tier, the destination of inbound demand $d \in D^I$ (or the origin of outbound demand $d \in D^O$) is not the customer $c \in C$ but a satellite. To avoid exclusive assignments as well, a set of potential satellites $Z(c) \subseteq Z$ is given with corresponding final distribution (or pickup) costs $k^Z(d, z, t)$ for demand $d$ from satellite $z$ in period $t$. Besides operating and handling costs also costs for disturbance through freight activities during the operation time of the service are included.

Moreover, each demand has a time-window: when it will be available at the origin $[a^o(d), b^o(d)]$, and when it can be delivered to the destination $[a^d(d), b^d(d)]$. If products are available at a certain time in an external zone, but do not have a deadline until they have to leave the depot, these bounds are set to infinity (or in the other case to minus infinity).

The travel-time between two points $i, j$ in the network is assumed to be time-dependent and defined by $\delta_{ij}(t)$. It is assumed that the travel times include the estimated congestion at departure time $t$ and are based on historical data. The service time for loading and unloading an urban-vehicle is $\delta(\tau)$.

### 3 Mixed-integer linear programming formulation

The goal is to select a set of urban vehicle services $R = \{r\}$ and the demand itinerary flows. Service $r$ starts at external zone $e(r)$, visits several satellites, and returns to the same external zone. The ordered set of visited satellites is given by $\sigma(r) = \{z_i \in Z, i = 1, \ldots, |\sigma(r)|\}$, such that if $r$ visits satellite $i$ before satellite $j$ then $i < j$. A vehicle of type $\tau(r)$, which is of mode $m(r)$, operates service $r$ and the associated costs are given by $k(r)$. The costs include again not only the operating costs of the route and unloading or loading activities, but also disturbance factors for freight activities during the operation time of the service.

The departure of the service at the origin $e(r)$ is $t(r)$. Then the urban vehicle arrives in period $t_1(r) = t(r) + \delta_{e(r)z_1(r)}(t(r))$ at the first satellite $z_1(r) \in \sigma(r)$. Since we also allow the waiting at specific satellites (for example of trams), the waiting time of service $r$ at satellite $z$ is denoted by $\delta(z, r)$. After performing loading and/or unloading operations and possible waiting, the urban vehicle leaves the first satellite at $t_1(r) + \delta(\tau(r)) + \delta(z_1(r), r)$. Thus the general schedule of service $r$ is given by the ordered set $\{t_i(r), i = 0, 1, \ldots, |\sigma(r)| + 1\}$, where $t_0(r) = t(r), t_i(r) = t_{i-1}(r) + \delta_{z_{i-1}(r)z_i(r)}(t_{i-1}(r)) + \delta(\tau(r)) + \delta(z_i(r), r)$, representing the period the service visits satellite $z_i \in \sigma(r)$, and the service finishes its route at the external $e(r)$ in period $t_{|\sigma(r)|+1}$. In the definition of $R$,
we account for the fact, that a tram track is used by public transportation or a satellite is not available in specific time slots.

To reflect the different compartments of the services, \( R^C(r) = \{ r^c \} \) defines the service of each compartment. For services with one compartment, obviously \(|R^C(r)| = 1\). If the service \( r \in R \) is operated, also all compartment services are operated. However, this differentiation is necessary to start in specific compartments with outbound demand while processing still inbound demand in other compartments.

Moreover, \( I = \{ i \} \) represents the set of all itineraries and \( I(d) \) the subset that may be used to satisfy customer demand \( d \in D \). Then, \( e(i) \) defines the used external zone for demand \( d \), \( z(i) \) the used satellite, and \( r(i) \) the operating service. For all inbound demands \( d \in D^I \), the external zones are the origin and the satellites the destination. For outbound demands \( d \in D^O \), it is the other way round. According to that separation \( I \) is divided into the inbound itineraries \( I^I \) and outbound itineraries \( I^O \). Since a compartment \( r^c \in R^C \) can only be loaded with outbound demand after all inbound demand is delivered, \( I^X(r^c) = \{(i_1, i_2) | i_1 \in I^I, i_2 \in I^O \} \), such that \( i_1 \) and \( i_2 \) cannot be combined \{ defines the set of forbidden itinerary combinations.

We use two sets of decision variables:

\[ \rho(r) = 1 \text{ if the urban vehicle service } r \in R \text{ is selected, } 0 \text{ otherwise; } \]

\[ \zeta(i) = 1 \text{ if itinerary } i \in I(d) \text{ of demand } d \in D \text{ is used, } 0 \text{ otherwise.} \]

The mixed-integer linear program to solve the first-tier service network design problem can then be written as

\[
\min \sum_{r \in R} k(r) \rho(r) + \sum_{d \in D} \sum_{i \in I(d)} k^Z(d, z, t) \zeta(i) + \sum_{d \in D} \sum_{i \in I(d)} k^E(d, e(i)) \zeta(i) \tag{1}
\]

subject to

\[
\sum_{i \in I(d)} \zeta(i) = 1 \quad \forall d \in D \tag{2}
\]

\[
\zeta(i_1) + \zeta(i_2) \leq 1 \quad \forall (i_1, i_2) \in I^X(r^c), r^c \in R^C(r), r \in R \tag{3}
\]

\[
\sum_{d \in D} \sum_{i \in I^I(d, r^c)} \text{vol}(d) \zeta(i) \leq u^e_c(r) \rho(r) \quad \forall r^c \in R^C(r), r \in R \tag{4}
\]

\[
\sum_{d \in D} \sum_{i \in I^O(d, r^c)} \text{vol}(d) \zeta(i) \leq u^e_c(r) \rho(r) \quad \forall r^c \in R^C(r), r \in R \tag{5}
\]

\[
\sum_{r \in R(t, e)} \rho(r) \leq n_{te} \quad \forall t \in T, e \in E, t = 1, \ldots, T \tag{6}
\]

\[
\sum_{t^- = t - \delta(t) + 1}^t \sum_{r \in R(z, t^-)} \rho(r) \leq u^z_{zt} \quad \forall z \in Z, t = 1, \ldots, T \tag{7}
\]

\[
\sum_{t^- = t - \delta(t) + 1}^t \sum_{r \in R(z, t^-)} \rho(r) \leq u^m_{zt} \quad \forall z \in Z, m \in M, t = 1, \ldots, T \tag{8}
\]
\[
\sum_{d \in D} \sum_{i \in I(d,z,t)} \text{vol}(d)\zeta(i) \leq u_{zt}^V \quad \forall z \in Z, t = 1, \ldots, T
\]

(9)

\[
\rho(r) \in \{0, 1\} \quad \forall r \in R
\]

(10)

\[
\zeta(i) \in \{0, 1\} \quad \forall i \in I
\]

(11)

The objective function (1) is minimizing the costs of selecting and operating a service, plus the costs for assigning a demand to a satellite, plus the costs for assign a demand to an external zone. The assignment costs include the operational costs at that terminal as well as the transportation costs from and to satellite/external zone.

Constraints (2) ensures that each demand is assigned to exactly one itinerary, and constraints (3) enforces that at maximum one itinerary is chosen, if both cannot be combined. This is the case if the pickup would start before the deliveries are finished in one compartment. Constraints (4) and (5) ensure that for in- and outbound demand the capacity of the urban vehicle compartments is not exceeded. In combination with the previous equation the capacity restriction is then ensured for the entire service. Constraints (6) ensure that the maximum number of available vehicles of one type in an external zone is never exceeded and constraints (7) and (8) limit the number of urban vehicles at a satellite in each period in total and per transportation mode. Finally, constraints (9) limit the maximum amount of demand which can be unloaded or loaded at a satellite in each period.

4 Conclusion

We will present a new model for transportation planning in CL. Compared to previous models, the developed model combines in- and outbound demand, considers different means of transport, and includes resource management. We are developing an exact solution method based on decomposition techniques to solve the problem efficiently. The full details of the model, method and numerical experiments will be presented at the conference. Other than the efficiency of the algorithm, we are interested in evaluating the value of integrating several modes and schedules, as well as the impact of explicitly considering outbound demand and management of resources supporting the CL system.

References


