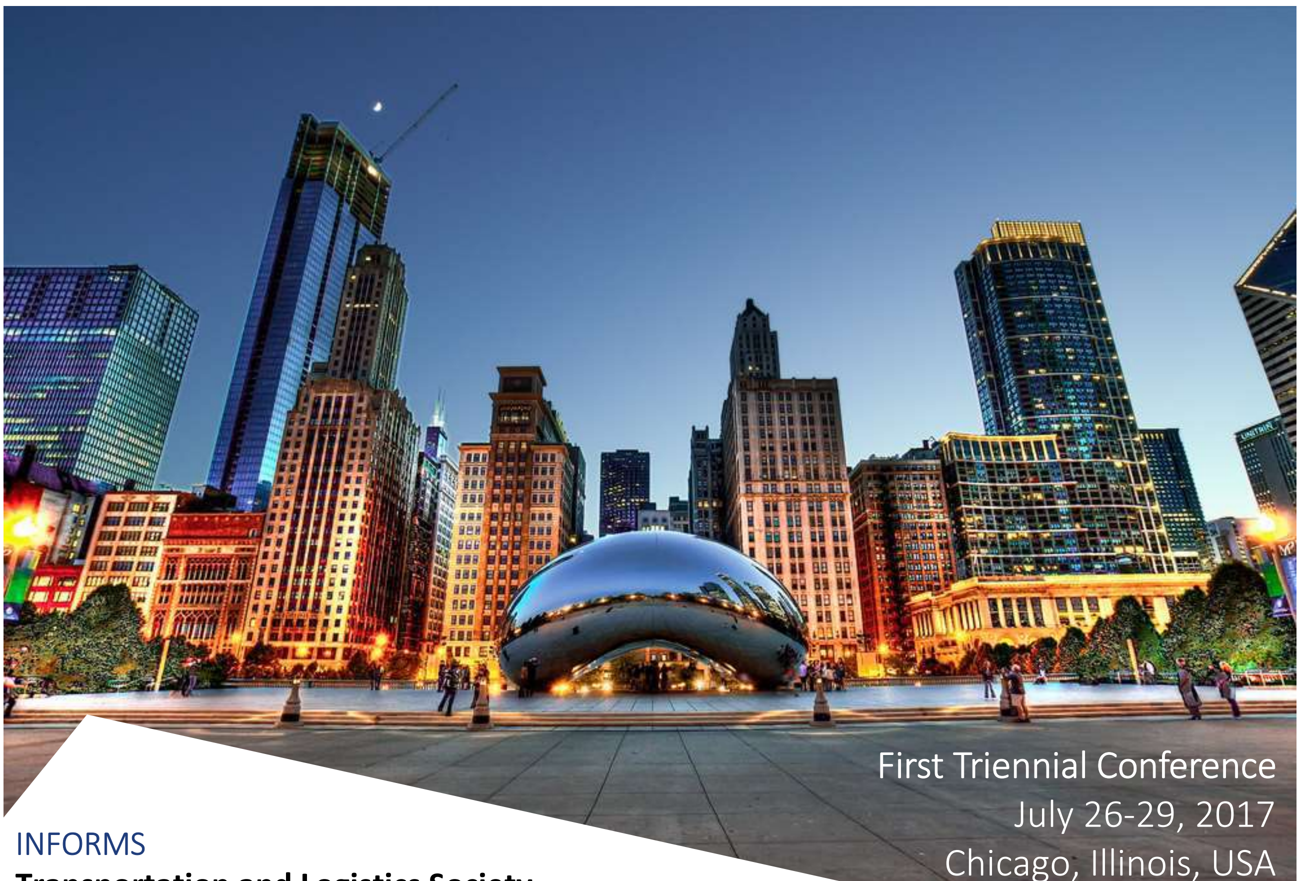


INFORMS
Transportation and Logistics Society

http://connect.informs.org/tsl/conferences/tsl-conference

Contents

Committees

Organizing Committee

Pitu Mirchandani – Chair
Arizona State University

Maciek Nowak – Local Chair
Loyola University Chicago

Mike Ball
University of Maryland

Mike Hewitt
Loyola University of Chicago

Warren Powell
Princeton University

Barry Thomas
University of Iowa

Scientific Committee

Nicole Adler, *Hebrew University of Jerusalem,* Israel
Niels Agatz, *Erasmus University,* Netherlands
Hillel Bar-Gera, *Ben Gurion University,* Israel
Rajan Batta, *University of Buffalo,* United States
Michel Bierlaire, *École Polytechnique Fédérale de Lausanne,* Switzerland
Steve Boyles, *University of Texas,* United States
Ann Campbell, *University of Iowa,* United States
Marielle Christiansen, *Norwegian University of Science and Technology*, Norway
Jean-Francois Cordeau, *HEC Montreal,* Canada
Francesco Corman, *TU Delft*, Netherlands
Theo Crainic, *Université du Québec à Montréal*, Canada
Mark Daskin, *University of Michigan*, United States
Maged Dessouky, *University of Southern California*, United States
Irina Dolinskaya, *National Science Foundation*, United States
Jan Ehmke, *Free University of Berlin*, Germany
Alan Erera, *Georgia Tech*, United States
Ozlem Ergun, *Northeastern University*, United States
Yueyue Fan, *UC Davis*, United States
Song Gao, *University of Massachusetts*, United States
Michel Gendreau, *École Polytechnique de Montréal*, Switzerland
Bernard Gendron, *Université de Montréal*, Canada
Monica Gentili, *University of Louisville*, United States
Ricardo Giesen, *Pontifical Catholic University Chile*, Chile
Bruce Golden, *University of Maryland*, United States
Kevin Gue, *University of Louisville*, United States
Geir Hasle, *SINTEF*, Norway
Andreas Hegyi, *TU Delft*, Netherlands
Mark Hickman, *University of Queensland*, Australia
Serge Hoogendoorn, *TU Delft*, Netherlands
Bogumil Kaminski, *Warsaw School of Economics*, Poland
Natalia Kliewer, *Free University of Berlin*, Germany
Martine Labbe, *Université libre de Bruxelles*, Belgium
Gilbert Laporte, *HEC Montreal*, Canada

Janny Leung, *Chinese University of Hong Kong*, Hong Kong
Ronghui Liu, *University of Leeds*, United Kingdom
Hong Lo, *Hong Kong University of Science and Technology*, Hong Kong
Yingyan Lou, *Arizona State University*, United States
David Lovell, *University of Maryland*, United States
Guglielmo Lulli, *Lancaster University*, United Kingdom
Hani Mahmassani*, Northwestern University*, United States
Lavanya Marla, *University of Illinois*, United States
Dirk Mattfeld, *Braunschweig University of Technology*, Germany
Elise Miller-Hooks, *University of Maryland*, United States
Juan Carlos Munoz, *Pontifical Catholic University Chile*, Chile
Marco Nie, *Northwestern University*, United States
Otto Anker Nielsen, *Technical University of Denmark*, Denmark
Amedeo Odoni, *MIT*, United States
Carolina Osorio, *MIT*, United States
Yanfeng Ouyang, *University of Illinois*, United States
Dario Pacciarelli, *Roma Tre University*, Italy
Dario Pacino, *Technical University of Denmark*, Denmark
Markos Papageorgiou, *Technical University of Crete*, Greece
Sophie Parragh, *University of Vienna*, Austria
Carlo Prato, *University of Queensland*, Australia
Harilaos Psaraftis, *Technical University of Denmark*, Denmark
Martin Savelsbergh, *Georgia Tech*, United States
Anita Schöbel, *University of Göttingen*, Germany
Karen Smilowitz, *Northwestern University*, United States
Larry Snyder, *Lehigh University*, United States
Senay Solak, *University of Massachusetts*, United States
Grazia Speranza, *University of Brescia*, United States
Alejandro Toriello, *Georgia Tech*, United States
Satish Ukkusuri, *Purdue University*, United States
Halit Uster, *Southern Methodist University*, United States
Hans Van Lint, *TU Delft*, United States
Tom Van Woensel, *Eindhoven University of Technology*, Netherlands
Francesco Viti, *University of Luxembourg*, Luxembourg
Stefan Voss, *University of Hamburg*, Germany
Thomas Vossen, *University of Colorado*, United States
Stein Wallace, *Norwegian School of Economics*, Norway
S. Travis Waller, *University of New South Wales*, Australia
Chip White, *Georgia Tech*, United States
Hai Yang, *Hong Kong University of Science and Technology*, Hong Kong
Yafeng Yin, *University of Florida*, United States
Lei Zhao, *Tsinghua University*, China
Xuesong Zhou, *Arizona State University*, United States

Venue

The conference will be held on Loyola's Water Tower Campus, located along Pearson Street, just off North Michigan Avenue, Chicago's famed "Magnificent Mile. The Water Tower Campus derives its name from the famous Chicago Water Tower, which survived the Great Chicago Fire in 1871. The campus sits in the shadow of the iconic John Hancock Center. Other nearby architectural landmarks are the Tribune Tower, the Wrigley Building, the Trump Tower, and the site of Fort Dearborn, around which the city of Chicago was founded. Holy Name Cathedral and the Roman Catholic Archdiocese of Chicago are located just south of campus, across Chicago Avenue. Cultural points of interest include the Museum of Contemporary Art, the Newberry Library, and Navy Pier. In addition to the unparalleled shopping on Michigan Avenue, the Water Tower Campus is also within a few minutes' walk of numerous dining and entertainment options.

Social Agenda

In addition to a broad look at the future of transportation, this workshop will provide numerous opportunities to network with colleagues and establish new working relationships.

A welcome reception overlooking downtown Chicago will start the event on Wednesday evening. The welcome reception is open to all registrants and has been generously sponsored by APICS.

APICS

APICS is the premier professional association for supply chain management and the leading provider of research, education and certification programs that elevate supply chain excellence, innovation and resilience. The APICS Certified in Production and Inventory Management (CPIM), APICS Certified Supply Chain Professional (CSCP), APICS Certified in Logistics, Transportation and Distribution (CLTD) and APICS Supply Chain Operations Reference Professional (SCOR-P) designations set the industry standard. With over 45,000 members, over 130,000 certified professionals, and approximately 300 channel partners internationally, APICS is transforming the way people do business, drive growth and reach global customers. APICS is dedicated to building greater awareness of the supply chain profession and develops lifelong learning content to ensure that the number and quality of students and professionals meets industry's needs. For more information, visit apics.org.

On Friday evening, we will adjourn for dinner at the Chicago Museum of Contemporary Art, with an opportunity to explore the collection prior to and after dinner. This reception is open to all those who submitted a regular (non-student) registration.

Lunches will be provided on site each day of the conference.

Plenary Sessions

Dr. Kimberly Ross

Transportation Optimization – Experiences from the Field

Thursday, July 27, 11:00 – 12:00 PM

Dr. Kimberly Ross has spent over 20 years consulting and developing optimization solutions for some of the largest shippers in North America including well-known retailers, grocers, food, and wholesale pharmaceutical distributors. In this talk, she will share some of her unique transportation modeling experiences including large-scale inbound planning for multi-model, dynamic cross-dock optimization typical of big-box retail, outbound planning with restricted time windows, split deliveries, tandem orchestration, and backhauls that are pervasive in the grocery market, and high density multi-stop routing optimization for food and pharmaceutical distribution with relays, cross-docks, and hybrid static/dynamic routing considerations. Additionally, she will share some of the challenges her customers face as they look to take advantage of promising supply chain optimization synergies as well as explore options that new technologies in big data and cloud computing may provide.

Dr. Kimberly Ross is Vice President of Research & Development at Manhattan Associates responsible for the Science team overseeing all optimization capabilities across the product suite including Transportation Management, Warehouse Management, Slotting Optimization, Demand Forecasting, and Inventory Optimization. She received her B.S. from Sanford University in Mathematical Sciences and her Ph.D. from Princeton University in Operations Research and has over 20 years of experience designing and implementing mathematical optimization algorithms to solve highly complex real-world problems, mostly in the transportation and logistics industries.

Dr. Peter Frazier

Providing Reliable Transportation at Uber

Friday, July 28, 11:00 – 12:00 PM

Ridesharing is revolutionizing transportation in cities. A central task in ridesharing is providing reliable transportation to riders and attractive earnings to drivers when neither group is under centralized control. This is especially challenging given that weather, traffic, sporting events, and holidays frequently cause hard-to-predict imbalances between riders' and drivers' willingness to participate in the market. We discuss approaches and mathematical models used at Uber to overcome this challenge, and provide an overview of other exciting new research questions in transportation opened by the growth of ridesharing.

Dr. Peter Frazier is an Associate Professor in the School of Operations Research and Information Engineering at Cornell University, and a Staff Data Scientist and Data Science Manager at Uber. He received a Ph.D. in Operation Research and Financial Engineering from Princeton University in 2009. His research is in optimal learning, sequential decision-making under uncertainty, and machine learning, focusing on applications in simulation, e-commerce, medicine, and biology. He is an associate editor for Operations research, ACM TOMACS, and IISE Transactions, and is the recipient of an AFOSR Young Investigator Award and an NSF CAREER Award.

Program Overview

Wednesday, July 26

3:15 – 5:15 PM	Connected Traffic & Transportation (Joint Session with ISTTT)
5:30 – 7:00 PM	Welcome reception

Thursday, July 27

7:30 – 6:00 PM	Registration
8:30 – 10:30 AM	Parallel sessions
10:30 – 11:00 AM	Break
11:00 – 12:00 PM	Plenary talk (Dr. Kimberly Ross)
12:00 – 1:00 PM	Lunch
1:00 – 2:30 PM	Parallel sessions
2:30 – 2:45 PM	Break
2:45 – 4:15 PM	Parallel sessions
4:15 – 4:30 PM	Break
4:30 – 6:00 PM	Parallel sessions

Friday, July 28

7:30 – 6:00 PM	Registration
8:30 – 10:30 AM	Parallel sessions
10:30 – 11:00 AM	Break
11:00 – 12:00 PM	Plenary talk (Dr. Peter Frazier)
12:00 – 1:00 PM	Lunch
1:00 – 2:30 PM	Parallel sessions
2:30 – 2:45 PM	Break
2:45 – 4:15 PM	Parallel sessions
4:15 – 4:30 PM	Break
4:30 – 6:00 PM	Parallel sessions

Saturday, July 29

9:00 – 11:00 AM	Parallel sessions
11:00 – 11:15 PM	Break
11:15 – 12:45 PM	Parallel sessions
1:00 – 1:30 PM	TSL Cross-region Dissertation Grant Winner

Technical Program

Connected Traffic & Transportation (Joint Session with ISTTT)

Location: Robert H. Lurie Building, 303 East Superior, Baldwin Auditorium
Time: 3:15 – 5:15 PM
Session Chair: Karen Smilowitz

Planning Reliable Service Facilities Under Continuous Traffic Equilibrium and Disruption Risks – *Zhaodong Wang, Yanfeng Ouyang**

Design and Control of Driverless Fleets of Electric Vehicles using Approximate Dynamic Programming – *Warren Powell*, Andy Deng, Lina Al Kanj, Alain Kornhauser*

From 'No Data' to 'Some Data' to 'Big Data' Towards a Cyber-Physical System for Proactive Traffic Management – *Pitu Mirchandani*, Kerem Demirtas, Viswanath Potluri*

Logistical Challenges at Mass Participation Events: Operations Research Models for Marathon Planning – *Karen Smilowitz*, Mehmet Basdere, Sanjay Mehrotra, George Chiampas, Jennifer Chan, Mike Nishi*

THURSDAY **8:30 – 10:30 AM**	MULTIMODAL TRANSPORTATION SERVICES Session TA1: **Air Traffic and Air Networks**	TRAFFIC & MOBILITY Session TA2: **Traffic Equilibrium**	VEHICLE ROUTING MODELS & APPLICATIONS Session TA3: **EV Charging Logistics**	SUPPLY CHAIN LOGISTICS & METHODS Session TA4: **Humanitarian Logistics**
Session Chair	**David Lovell**	**Carlo Prato**	**Halit Uster**	**Lavanya Marla**
8:30 AM	Choice-Based Airline Fleet Assignment – *Chiwei Yan*, Cynthia Barnhart, Vikrant Vaze*	Analyzing Departure Time Choice in a Bottleneck with Stochastic Service Time – *Gege Jiang*, Hong Lo*	Locating Refueling Points on Lines and Comb-Trees – *Pitu Mirchandani, Yazhu Song**	Using Drones to Minimize Latency in Distribution Systems – *Mohammad Moshref-Javadi*, Seokcheon Lee*
9:00 AM	The Most Reliable Path Problem for Airline Travel with Connections – *Michael Redmond*, Ann Campbell, Jan Ehmke*	An Algorithm for Transit Assignment Problem with Flow-Dependent Dwell Times – *Yufeng Zhang*, Alireza Khani*	Modeling Electric Vehicle Charging Demand – *Guus Berkelmans, Wouter Berkelmans, Nanda Piersma, Rob van der Mei, Elenna Dugundji**	A Novel Formulation and a Column Generation Technique for a Rich Humanitarian Logistic Problem – *Ohad Eisenhandler, Michal Tzur**
9:30 AM	Modeling Flight Delay Propagation: A New Analytical-Econometric Approach – *Nabin Kafle, Bo Zou**	Statistical Inference of Probabilistic Origin-destination Demand Using Day-To-Day Traffic Data – *Wei Ma*, Zhen Qian*	Electric Vehicle Routing with Uncertain Charging Station Availability & Dynamic Decision Making – *Nicholas Kullman*, Justin Goodson, Jorge Mendoza*	Humanitarian Medical Supply Chain in Disaster Response: Role and Challenges – *Irina Dolinskaya*, Maria Besiou, Sara Guerrero-Garcia*
10:00 AM	Sources of Flight Inefficiency in the National Airspace System: An Econometric Approach – *Mark Hansen*, Yulin Liu, Michael Ball, David Lovell, Cara Chuang*	Threshold-Based Stochastic User Equilibrium Models – *David Watling, Thomas Rasmussen, Otto Nielsen, Carlo Prato**	Network Design for In-Motion Wireless Charging of Electric Vehicles in Urban Traffic Networks – *Mamdouh Mubarak*, Halit Uster, Khaled Abdelghany, Mohammad Khodayar*	Cooperative Humanitarian Logistics Models for Highly Resource-Constrained Settings – *Lavanya Marla**

THURSDAY 1:00 – 2:30 PM	MULTIMODAL TRANSPORTATION SERVICES	TRAFFIC & MOBILITY	VEHICLE ROUTING MODELS & APPLICATIONS	SUPPLY CHAIN LOGISTICS & METHODS
	Session TB1: **IP Methods in Air Traffic Control**	Session TB2: **Pricing and Competition**	Session TB3: **VRP Exact Methods**	Session TB4: **Games and Collaboration**
Session Chair	**Michael Ball**	**Song Gao**	**Michel Gendreau**	**Nicole Adler**
1:00 PM	Optimizing the Slot Allocation on a Network of Airports – *Paola Pellegrini*, Tatjana Bolić, Lorenzo Castelli, Raffaele Pesenti*	Competitive Rebalancing in One-Way Car-Sharing – *Szymon Albinski*, Stefan Minner*	An Integer Programming Approach for the Time-Dependent Traveling Salesman Problem with Time Windows – *Agustin Montero, Isabel Mendez-Diaz, Juan Jose Miranda Bront**	Multi-Round Combinatorial Auctions for Carrier Collaboration – *Margaretha Gansterer*, Richard Hartl, Martin Savelsbergh*
1:30 PM	Greedy Policies for a Dynamic Stochastic Transportation Problem, and an Application to Air Traffic Management – *Alexander Estes*, Michael Ball*	The Costs and Benefits of Ridesharing: Sequential Individual Rationality and Sequential Fairness – *Ragavendran Gopalakrishnan*, Koyel Mukherjee, Theja Tulabandhula*	A Mixed-Integer Linear Program for the Traveling Salesman Problem with Structured Time Windows – *Philipp Hungerlaender, Christian Truden**	Collaborative Vehicle Routing with Excess Vehicle Capacity in Urban Last-Mile Deliveries – *Joydeep Paul*, Niels Agatz, Remy Spliet, René De Koster*
2:00 PM	A Mechanism for Auctioning Airport Landing Slots with Explicit Valuation of Congestion – *Michael Ball*, Alex Estes, Mark Hansen, Yulin Liu*	Cooperative Scheme - An Alternative Approach to Equitable and Pareto-Improving System Optimum – *Sayeeda Ayaz*, Song Gao, Hyoshin Park*	A Branch-and-Price Algorithms for a Multi-Attribute Technician Routing and Scheduling Problem – *Michel Gendreau*, Ines Mathlouthi, Jean-Yves Potvin*	Competition in Congested Service Networks: The Case of Air Traffic Control Provision in Europe – *Nicole Adler*, Eran Hanany, Stef Proost*

THURSDAY **2:45 – 4:15 PM**	MULTIMODAL TRANSPORTATION SERVICES Session TC1: **Airport Operations**	TRAFFIC & MOBILITY Session TC2: **Pricing Shared Mobility**	VEHICLE ROUTING MODELS & APPLICATIONS Session TC3: **VRP Extensions**	SUPPLY CHAIN LOGISTICS & METHODS Session TC4: **Delivery Service Network**
Session Chair	**Senay Solak**	**Tunay Tunca**	**Daniele Vigo**	**Niels Agatz**
2:45 PM	Efficiency, Equity and On-Time Performance Objectives in Airport Demand Management – *Alexandre Jacquillat*, Vikrant Vaze*	Split the Bill by Sharing the Ride-Sharing Services: A Study on Optimal Ride-Sharing Pricing – *Jagan Jacob*, Ricky Roet-Green*	Multi-Modal Variations of the Vehicle Routing Problem – *Marc-Antoine Coindreau, Olivier Gallay*, Zufferey Nicolas*	An On-Demand Same-Day Delivery Service Using Direct Peer-to-Peer Transshipment Strategies – *Wei Zhou, Jane Lin**
3:15 PM	A Data-Splitting Algorithm for Flight Sequencing and Scheduling on Two Runways – *Rakesh Prakash*, Jitamitra Desai*	Inventory Rebalancing and Minimum Stop-Over Routes for One-Way Electric Vehicle Sharing Systems – *Yinglei Li*, Sung Hoon Chung*	Multi-Commodity Two-Echelon Vehicle Routing Problem with Time Windows – *Tom Van Woensel**	Balancing Availability and Profitability in E-Fulfillment with Revenue Management and Predictive Routing – *Catherine Cleophas*, Jan Fabian Ehmke, Charlotte Köhler, Magdalena Lang*
3:45 PM	Lower Cost Departures for Airlines: Optimal Gate and Metering Area Allocation Policies Under Departure Metering Concept – *Heng Chen*, Senay Solak*	An Empirical Analysis of Price Formation, Utilization, and Value Generation in Ride Sharing Services – *Liu Ming, Tunay Tunca*, Yi Xu, Weiming Zhu*	The Vehicle Routing Problem with Private Fleet and Common Carrier Extension and Exact Algorithm – *Said Dabia, David Lai, Daniele Vigo**	Heuristic Approaches to the Same-Day Delivery Problem – *Alp Arslan*, Niels Agatz, Rob Zuidwijk*

THURSDAY **4:30 – 6:00 PM**	MULTIMODAL TRANSPORTATION SERVICES Session TD1: **Bus and Taxi Transport**	TRAFFIC & MOBILITY Session TD2: **Optimization in Shared Mobility**	VEHICLE ROUTING MODELS & APPLICATIONS Session TD3: **Dynamic Routing**	SUPPLY CHAIN LOGISTICS & METHODS Session TD4: **Service Network Design**
Session Chair	**Mark Hickman**	**Carolina Osorio**	**Barrett Thomas**	**Mark Hewitt**
4:30 PM	Choice between Metro and Taxi under Travel Time Variability – *Gege Jiang*, Hong Lo*	Optimizing the Profitability and Quality of Service in Carshare Systems Under Demand Uncertainty – *Mengshi Lu*, Siqian Shen, Zhihao Chen*	Route-Based Markov Decision Processes for Dynamic Vehicle Routing Problems – *Justin Goodson*, Marlin Ulmer, Dirk Mattfeld, Barrett Thomas*	A Benders' Decomposition Approach for Airline Timetable Development and Fleet Assignment – *Keji Wei*, Vikrant Vaze*
5:00 PM	Multi-Cycle Optimal Taxi Routing with E-Hailing – *Xinlian Yu*, Song Gao, Hyoshin Park, Xianbiao Hu*	Ridesharing in a Mobility-On-Demand System – *Samitha Samaranayake*, Harshith Guntha, Kevin Spieser, Emilio Frazzoli*	Scalable Anticipatory Policies for the Dynamic and Stochastic Pickup and Delivery Problem – *Gianpaolo Ghiani, Emanuele Manni*, Alessandro Romano*	Service Network Design of Bike Sharing Systems: Formulation and Solution Method – *Bruno Albert Neumann Saavedra*, Dirk Mattfeld, Teodor Gabriel Crainic, Bernard Gendron, Michael Römer*
5:30 PM	Machine Learning Methods to Predict Bus Travel Speeds and Analysis of the Impact of Different Predictive Variables – *Jan Berczely, Ricardo Giesen**	A Discrete Simulation-Based Optimization Algorithm for Two-Way Car-Sharing Network Design – *Tianli Zhou*, Carolina Osorio, Evan Fields*	Enough Waiting for the Cable Guy - Estimating Arrival Times for Service Vehicle Routing – *Barrett Thomas*, Marlin Ulmer*	Enhanced Dynamic Discretization Discovery Algorithms for Service Network Design Problems – *Mike Hewitt**

FRIDAY **8:30 – 10:30 AM**	MULTIMODAL TRANSPORTATION SERVICES Session FA1: **Transit Fare and Revenue**	TRAFFIC & MOBILITY Session FA2: **Traffic Management**	VEHICLE ROUTING MODELS & APPLICATIONS Session FA3: **Dynamic Pick Up and Delivery**	SUPPLY CHAIN LOGISTICS & METHODS Session FA4: **Location Routing**
Session Chair	**Hai Yang**	**Lili Du**	**Martin Savelsbergh**	**Francesco Viti**
8:30 AM	Ticket Portfolio Planning for Urban and Regional Public Transport Networks – *Jörn Schönberger**	Morning Commute Problem with Queue-Length-Dependent Bottleneck Capacity – *Jin-Yong Chen, Rui Jiang*, Xin-Gang Li, Mao-Bin Hu, Bin Jia*	Dynamic Pickup and Delivery Problems with Transfers – *Afonso Sampaio, Martin Savelsbergh, Lucas Veelenturf*, Tom Van Woensel*	Finding Optimal Park-and-Ride Facility Locations in an Urban Network – *Pramesh Kumar*, Alireza Khani*
9:00 AM	Optimal Discount Policies for Transit Agencies: The Case of Pass Programs and Loyalty Programs – *Mehdi Nourinejad*, Amir Gandomi, Joseph Y. J. Chow, Matthew J. Rooda*	Modelling User Behavior at a Stochastic Bottleneck – *Daphne van Leeuwen*, Peter van de Ven*	How Many Vehicles Do We Need? Pickup and Delivery Problem with Synchronized Tasks and Transfers – *Monirehalsadat Mahmoudi*, Junhua Chen, Xuesong Zhou*	Reliable Facility Location Design with Imperfect Information: Continuum and Discrete Models – *Lifen Yun, Hongqiang Fan, Xiaopeng Li**
9:30 AM	Pricing for a Last Mile Transportation System – *Hai Wang*, Yiwei Chen*	Recasting Intersection Automation as a Connected-and-Automated-Vehicle (CAV) Scheduling Problem within Heterogeneous Traffic Environment – *Pengfei (Taylor) Li*, Xuesong Zhou*	Order Acceptance Mechanisms for Same-Day Delivery – *Mathias Klapp*, Alan Erera, Alejandro Toriello*	Location-Routing Problems with Economies of Scale – *James Bookbinder*, Xiaoyang Pi*
10:00 AM	A Fare-Reward Scheme for Commuters in Transit Bottleneck – *Yili Tang*, Hai Yang*	Distributed Computation Based Constrained Model Predictive Control for a Mixed Flow Platoon – *Siyuan Gong*, Lili Du*	Optimization Algorithms for Meal Delivery Operations – *Damian Reyes*, Alan Erera, Martin Savelsbergh*	Exact and Approximate Optimal Route Set Generation in Sensor Location Problems – *Marco Rinaldi*, Francesco Viti*

FRIDAY 1:00 – 2:30 PM	MULTIMODAL TRANSPORTATION SERVICES Session FB1: **Transit Networks**	TRAFFIC & MOBILITY Session FB2: **Network Design and Location Decisions**	VEHICLE ROUTING MODELS & APPLICATIONS Session FB3: **Stochastic Routing**	SUPPLY CHAIN LOGISTICS & METHODS Session FB4: **Freight Transportation**
Session Chair	**Juan Carlos Munoz**	**Yanfeng Ouyang**	**Stefan Minner**	**Ronald Askin**
1:00 PM	An Internal Bounding Method for Line and Shuttle Bus Planning – *Evelien van der Hurk**	Emission-controlled Pavement Management Scheduling – *Umit Tursun*, Hasan Ozer*	Vehicle Routing with Space- and Time-Dependent Stochastic Travel Times – *Stein W. Wallace*, Zhaoxia Guo, Michal Kaut*	Scheduled Service Network Design and Revenue Management with an Intermodal Barge Transportation Illustration – *Teodor Gabriel Crainic*, Ioana Bilegan, Yunfei Wang*
1:30 PM	Urban Transit Network Design and Timetabling Problem for Multi-Depot Round-Trip Routes – *James Chu**	Scheduling Work Zones in Transportation Service Networks – *Dening Peng*, Pitu Mirchandani*	A Two-Phase Safe Vehicle Routing and Scheduling Problem: Formulations and Solution Algorithms – *Aschkan Omidvar*, Eren Ozguven, Arda Vanli, Reza Tavakkoli-Moghaddam*	Load Commitment Policies for the Stochastic Advance Booking Problem for Truckload Trucking – *Juliana Nascimento*, Hugo Simao, Warren Powell*
2:00 PM	The Limited-Stop Bus Service Design Problem with Stochastic Passenger Assignment – *Guillermo Soto, Homero Larrain, Juan Carlos Munoz**	Reliable Sensor Location for Object Positioning and Surveillance via Trilateration – *Kun An, Siyang Xie, Yanfeng Ouyang*	Optimal A-Priori Tour and Restocking Policies for the Vehicle Routing Problem with Stochastic Demands – *Alexandre Florio, Richard F. Hartl, Stefan Minner**	Logistic Network Design for Daily Cyclic Truck Routes – *Ronald Askin*, Zhengyang Hu, Guiping Hu*

FRIDAY **2:45 – 4:15 PM**	MULTIMODAL TRANSPORTATION SERVICES Session FC1: **Containers Logistics**	TRAFFIC & MOBILITY Session FC2: **Traffic Flow Modeling**	VEHICLE ROUTING MODELS & APPLICATIONS Session FC3: **Statistical Data Analytics for Routing and Location**	SUPPLY CHAIN LOGISTICS & METHODS Session FC4: **Movement and Inventory Control in Supply Chains**
Session Chair	**Patrick Jaillet**	**Hillel Bar-Gera**	**Bruce Golden**	**Rajan Batta**
2:45 PM	Balancing the Trade-Off in Route Choice and Demurrage Costs in Inland Container Logistics – *Bernard Zweers*, Rob van der Mei, Sandjai Bhulai*	The Role of Stochasticity in Traffic Flow Instabilities – *Junfang Tian*, Rui Jiang, Martin Treiber*	A Novel Statistical Algorithm for Very Large-scale Vehicle Routing Problems with Time Windows – *Mayank Baranwal, Lavanya Marla*, Srinivasa Salapaka, Carolyn Beck*	Dynamic Capacity Logistics and Inventory Control – *Satya Malladi*, Alan Erera, Chelsea White III*
3:15 PM	An Integrated Model for Inbound Train Split and Container Loading in an Intermodal Railway Terminal – *Bruno Bruck*, Jean-François Cordeau, Emma Frejinger*	A Kalman Filter Approach for Dynamic Calibration of A Simplified Lower-Order Car Following Model – *Kerem Demirtas*, Pitu Mirchandani, Xuesong Zhou*	Facility Location and Design Decisions from Public Data – *Kalyan Talluri, Muge Tekin**	Planning the Fuel Supply to Gas Stations According to the Concept of Carrier-Mannaged Inventory - an Optimization Approach – *Paweł Hanczar**
3:45 PM	The Stochastic Container Relocation Problem – *Virgile Galle*, Setareh Borjian Boroujeni, Vahideh Manshadi, Cynthia Barnhart, Patrick Jaillet*	Representation Requirements for Perfect First-In-First-Out Verification in Continuous Flow Dynamic Models – *Hillel Bar-Gera*, Malachy Carey*	Addressing Uncertainty in Meter Reading for Utility Companies Using RFID Technology – *Debdatta Sinha Roy*, Bruce Golden, Edward Wasil*	Exploration of Strategies to Form Convoys to Facilitate Effective Movement of Items – *Rajan Batta*, Azar Sadeghnejad Barkousaraie, Moises Sudit*

FRIDAY **4:30 – 6:00 PM**	MULTIMODAL TRANSPORTATION SERVICES Session FD1: **Maritime Shipping and Fleets**	TRAFFIC & MOBILITY Session FD2: **Ride Sharing and Parking**	VEHICLE ROUTING MODELS & APPLICATIONS Session FD3: **Routing Applications**	SUPPLY CHAIN LOGISTICS & METHODS Session FD4: **Risk Analysis in Network Transportation**
Session Chair	**Stein W. Wallace**	**Yafeng Yin**	**Burcu Keskin**	**Laura Wynter**
4:30 PM	Speed Optimization Across Different Emission Control Zones – *Line Reinhardt*, Christos Kontovas*	Optimal Two-Sided Pricing Strategies of Shared E-Parking Platform with Elastic Demand – *Chaoyi Shao*, Hai Yang, Fangni Zhang*	Heuristics and Lower Bounds for Robust Heterogeneous Vehicle Routing Problems Under Demand Uncertainty – *Anirudh Subramanyam*, Panagiotis Repoussis, Chrysanthos Gounaris*	Hazardous-Materials Network Design Problem with Behavioral Conditional Value-at-Risk – *Liu Su*, Changhyun Kwon*
5:00 PM	A Column-Row-Generation Approach to Liner Shipping Network Design – *Jun Xia, Zhou Xu**	Integration of an Aggregated Dynamic Traffic Model with Advanced Optimization Techniques for Strategic Transit-Parking Planning – *Joana Cavadas*, António Pais Antunes, Nikolas Geroliminis*	Optimal Snow Plow Routing with Route Continuity Constraint – *Luning Zhang*, Jing Dong*	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*
5:30 PM	Planning for Charters: A Stochastic Maritime Fleet Composition and Deployment Problem – *Xin Wang*, Kjetil Fagerholt, Stein W. Wallace*	Parking Provision for Ride-Sourcing Services – *Zhengtian Xu*, Yafeng Yin, Liteng Zha*	Patrol Routing with Hybrid Electric Vehicles – *Mesut Yavuz*, Burcu Keskin, Huseyin Ergin*	Entity Resolution and Vessel Modeling for Maritime Situational Awareness – *Shiau Hong Lim*, Yeow Khiang Chia, Laura Wynter*

SATURDAY 9:00 – 11:00 AM	Multimodal Transportation Services Session SA1: **Rail Transport**	Traffic & Mobility Session SA2: **Simulation-Based Analysis and Optimization**	Vehicle Routing Models & Applications Session SA3: **VRP in Services**	Supply Chain Logistics & Methods Session SA4: **Behavioral Data and Demand Estimation**
Session Chair	**Dario Pacciarelli**	**Ronghui Liu**	**Lei Zhao**	**Yueyue Fan**
9:00 AM	The Load Planning Problem for Double-Stack Trains at Intermodal Terminals – *Serena Mantovani*, Gianluca Morganti, Nitish Umang, Teodor Gabriel Crainic, Emma Frejinger, Eric Larsen*	Traffic Management Strategies for Trucks in Urban Environments Based on a Fast Traffic Simulation Algorithm – *Michele Simoni*, Christian Claudel*	Self-Sustained Car-And-Ride Sharing Design and Optimization for Improving the Mobility of Underserved Communities – *Miao Yu*, Siqian Shen*	Household Use of Autonomous Vehicles: Modeling Framework and Traveler Adaptation – *Yashar Khayati, Jee Eun Kang*, Mark Karwan, Chase Murray*
9:30 AM	Tactical Block and Car Planning for Intermodal Trains – *Gianluca Morganti*, Teodor Gabriel Crainic, Emma Frejinger, Nicoletta Ricciardi*	Simulation Based Quantification of the Potential Impacts of Incidents on Connected Vehicle Applications – *Abdullah Kurkcu*, Fan Zuo, Jingqin Gao, Kaan Ozbay*	The Multi-Period Service Planning and Routing Problem – *Albert H. Schrotenboer*, Evrim Ursavas, Iris F.A. Vis*	Estimating Primary Demand of One-Way Vehicle Sharing Systems – *Chiwei Yan*, Chong Yang Goh*
10:00 AM	Optimization of Handouts for Rolling Stock Rotations Visualization – *Boris Grimm*, Ralf Borndoerfer, Thomas Schlechte, Markus Reuther*	An Efficient Sampling Method for Stochastic Simulation-based Transportation Optimization – *Timothy Tay*, Carolina Osorio*	Covering Tour Problem with an Application to School Bus Routing: Analysis of Single Vehicle Tours on a Grid – *Liwer Zeng*, Sunil Chopra, Karen Smilowitz*	Modeling the Acceptability of Crowdsourced Goods Deliveries – *Aymeric Punel, Alireza Ermagun, Amanda Stathopoulos**
10:30 AM	Real-Time Near-Optimal Train Scheduling and Routing in Complex Railway Networks – *Marcella Samà, Andrea D'Ariano, Dario Pacciarelli*, Francesco Corman*	A Within-Day Microscopic Dynamical Model of Route Choice and Responsive Traffic Signal Control – *Ronghui Liu*, Mike Smith*	Coordinated Delivery to Nanostores in Megacities – *Ruidian Song, Lei Zhao*, Jan C. Fransoo, Tom Van Woensel*	Travel Demand Estimation Using Heterogeneous Data Pieces: Addressing Stochasticity and Observability Issues – *Yudi Yang, Yueyue Fan*, Roger Wets*

SATURDAY **11:15 – 12:45 PM**	MULTIMODAL TRANSPORTATION SERVICES Session SB1: **Multiobjective and Multidimensional Logistics**	TRAFFIC & MOBILITY Session SB2: **Traffic Data Analysis**	VEHICLE ROUTING MODELS & APPLICATIONS Session SB3: **VRP in Health and Food Delivery**	SUPPLY CHAIN LOGISTICS & METHODS Session SB4: **City Logistics and Inventory Control**
Session Chair	**Fabien Lehuédé**	**Serge Hoogendoorn**	**Ann Campbell**	**Teodor Gabriel Crainic**
11:15 AM	Multi-Criteria Decision Making when Planning & Designing Sustainable Multi-Modal Transportation in a Corridor – *Marie Louis*, Eric Gonzales*	A Data-Driven and Integrated Evaluation of Area-Wide Impacts of Double Parking Using Macroscopic and Microscopic Models – *Jingqin Gao, Kaan Ozbay*, Michael Marsico*	An ALNS For a Rich Home Health Care Routing and Scheduling Problem – *Florian Grenouilleau*, Antoine Legrain, Nadia Lahrichi, Louis-Martin Rousseau*	Value Function Approximation-based Dynamic Look-ahead Policies for Stochastic-Dynamic Inventory Routing in Bike Sharing Systems – *Jan Brinkmann, Marlin Ulmer, Dirk Mattfeld**
11:45 AM	Bin-Packing Problems with Load Balancing and Stability Constraints – *Alessio Trivella*, David Pisinger*	Urban Trajectory Analytics: Day-Of-Week Movement Pattern Mining Using Tensor Factorization – *Jiwon Kim*, Kianoosh Soltani Naveh*	Team Orienteering with Uncertain Rewards and Service Times with an Application to Phlebotomist Intra-Hospital Routing – *Huan Jin*, Barrett Thomas*	Synchronizing City Logistics with Sliding Time Windows – *Saijun Shao*, Gangyan Xu, Ming Li, George Q. Huang*
12:15 PM	Multi-Directional Local Search for A Bi-Objective Vehicle Routing Problem with Lexicographic Minimax Load Balancing – *Fabien Lehuédé*, Olivier Péton, Fabien Tricoire*	Opportunities for Floating Car Data in Integrated Traffic Management: The Case of Queue Estimation – *Serge Hoogendoorn*, Erik-Sander Smits, Jaap Van Kooten*	The Restaurant Delivery Problem – *Marlin Ulmer*, Barrett Thomas, Ann Campbell, Nicholas Woyak*	Multi-Modal Scheduled Service Network Design for Two-Tier City Logistics System with Resource Management – *Pirmin Fontaine*, Teodor Gabriel Crainic, Ola Jabali, Walter Rei*

TSL CROSS-REGION DISSERTATION GRANT WINNER

Saturday, 1:00 – 1:30 PM
Maciek Nowak

Multi-objective stochastic optimization models for managing a bike sharing system – *Rossana Cavagnini*, Luca Bertazzi, Francesca Maggioni, Mike Hewitt*

Special Joint Session with ISTTT

Wednesday 3:15 – 5:15 PM Robert H. Lurie Building, 303 East Superior, Baldwin Auditorium

Session Chair: Karen Smilowitz

3:15 Planning Reliable Service Facilities Under Continuous Traffic Equilibrium and Disruption Risks
*Zhaodong Wang, Yanfeng Ouyang**
University of Illinois Urbana-Champaign

3:45 Design and Control of Driverless Fleets of Electric Vehicles using Approximate Dynamic Programming
Warren Powell, Andy Deng, Lina Al Kanj, Alain Kornhauser*
Princeton University

4:15 From 'No Data' to 'Some Data' to 'Big Data' Towards a Cyber-Physical System for Proactive Traffic Management
Pitu Mirchandani, Kerem Demirtas, Viswanath Potluri*
Arizona State University

4:45 Logistical Challenges at Mass Participation Events: Operations Research Models for Marathon Planning
[1]Karen Smilowitz, [1]Mehmet Basdere, [1]Sanjay Mehrotra, [2]George Chiampas, [1]Jennifer Chan, [3]Mike Nishi*
[1]Northwestern University, [2]Northwestern and Bank of America Chicago Marathon and Shamrock Shuffle, [3]Chicago Event Management

Design and Control of Driverless Fleets of

Electric Vehicles using Approximate Dynamic Programming

Warren B. Powell, Andy Deng, Lina al-Kanj, Alain Kornhauser

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

December 12, 2016

By year 2020 or 2021, almost every major auto company, along with fleet operators such as Uber and Lyft, have announced plans to put driverless vehicles on the road. At the same time, electric vehicles are quickly emerging as a next-generation technology that is cost effective, in addition to offering the benefits of reducing the carbon footprint by allowing cars to be charged with the most effective generators over the grid.

The combination of a centrally managed fleet of driverless vehicles, along with the operating characteristics of electric vehicles, is creating a transformative new technology that offers significant cost savings while still offering the same or higher levels of service. Some examples of the benefits are: 1) lower operating costs, 2) lower capital costs due to higher utilization and equipment that is better tuned to particular uses (e.g. smaller cars for commuting), 3) dramatic reduction in insurance due to lower accident rates and 4) significant reduction in parking facilities.

The single biggest cost of an electric vehicle today is the battery, which has to be sized to meet the requirements of the longest trip that a driver *might* make. This requirement vanishes, however, in a fleet setting where a fleet operator can dynamically handle the recharging of a vehicle. The fleet operator will understand the trip characteristics for each zone and can ensure that the cars in the area are sufficiently charged to meet these needs. Customers with unexpectedly long trips can be met by searching around the fleet for a car with a larger battery, and a sufficiently large charge.

Currently, ride sharing operators are not concerned with a driver after a match has been made. The situation changes when the fleet operator is responsible for the car. If the car had a driver, then the company might be concerned with where the driver lives and how far the trip takes the driver away from his/her home. This is the issue that has to be managed in truckload trucking, where drivers represent an unusually complex resource (see Simao et al. (2009) for a complete description). For an electric vehicle, the major challenge is recharging the battery. The decision of when to recharge needs to take into consideration the nature of trips from locations close to where the car is now, the possibility of repositioning the car to areas with a higher density of trips, and the anticipated number of trips compared to the cars available over the day (afternoon peaks are a real challenge).

Ultimately, a major question is the design of the fleet itself. While an individual driver owning a vehicle has to consider the distribution of trip lengths he/she may have to make with a car, a fleet operator simply needs to ensure that there are enough cars operating with sufficient charge in their batteries to meet the needs at a point in time. Of course, if the batteries are too small, cars will spend too much time moving to and from charging stations. In addition, there will be a strong incentive to have cars that can get through a peak period without charging at all, since this helps to determine the minimum fleet size. Even with these concerns, a fleet operator can likely operate with much smaller batteries than would be required for a single self-owned car.

We anticipate that a fleet operator will likely want to have a distribution of battery sizes, in addition to offering cars that hold different numbers of passengers (as Uber and Lyft already do). However, we anticipate that the average battery size is likely to be much smaller, introducing yet another potential source of cost savings.

This research has the following goals:

- We propose to use approximate dynamic programming to develop high-quality operational control strategies to determine which car (given the current charge level) is best for a particular trip (considering the length of the trip and the destination), when a car should be recharged, and when it should be repositioned to a different zone which offers a higher density of trips.
- We then propose to use outputs (in the form of value functions) from the operational planning model to optimize the distribution of battery capacities in the fleet.
- We wish to determine the number of cars required to provide a high level of service, and from this to understand the economics of a driverless fleet of electric vehicles.

Our work is going to build on two successful applications of approximate dynamic programming for fleet management. The first and most relevant to this study is the work presented in Simao et al. (2009) for truckload trucking, which is the application closest to this project. Both the truckload example and the driverless fleets involve the dynamic assignment of resources (trucks, cars) to tasks (loads, trips), where a major issue is determining which resource to assign to each task.

One issue missing from the truckload example is empty repositioning. In Simao et al. (2009), each decision involves assigning a driver to a load. At no time does the company simply move the driver to a location where a load might arise, an issue that was prominent in early work (Powell (1987), Powell et al. (1988), Cheung & Powell (1996)); see Crainic (2003) for a thorough review. We anticipate, however, that we will need to move drivers from one location (e.g. a suburban neighborhood) back to locations where there is a higher density of traffic. Since customers need to be served quickly, it will not always be possible to wait until a call comes in and then move the car. This means moving a car in anticipation of a demand that might arise.

The problem of moving transportation resources in anticipation of uncertainty demands has a long history in the transportation literature, beginning with applications motivated by the freight car distribution problem in railroads (Jordan & Turnquist (1983) was the earliest work to address the problem). This early work motivated a stream of papers on the "dynamic vehicle allocation problem" (Cheung & Powell (1996), Powell & Godfrey (2002), Powell & Topaloglu (2005), Topaloglu & Powell (2006)) which focused on dynamic programming approximations. Our work here builds on the work of Cheung & Powell (1996) and Topaloglu & Powell (2006), but is adapted to handle the dimension of representing the charge level in a car. For example, we need this logic to optimize the charge process so that the fleet can get through daily peaks while minimizing the number of cars being charged during the peak.

1 Model

We build on the modeling frameworks used in Simao et al. (2009) and Bouzaiene-Ayari et al. (2014). There are five components to any sequential dynamic model:

State variables:
We model our resources (the cars) using

a = The attributes of a car, which include the current location, the estimated time of arrival (if it is moving to the location), and the charge level of the battery (or estimated charge at the arrival of the destination), with $a \in \mathcal{A}$,

R_{ta} = The number of cars with attribute a at time t; $R_t = (R_{ta})_{a \in \mathcal{A}}$.

We model the demands using

b = The attributes of a trip (origin, destination, pickup time), with $b \in \mathcal{B}$,

D_{tb} = The number of trips with attribute b in the system at time t; $D_t = (D_{tb})_{b \in \mathcal{B}}$.

We model our state variable as $S_t = (R_t, D_t)$.

Decision variables:
We let $\mathcal{D}_t$ be the different types of decisions, which are made up of the following sets:

$$\begin{aligned}
\mathcal{D}^s &= \text{Set of charging stations, where } d \in \mathcal{D}^s \text{ means "move to a particular charging station,"} \\
\mathcal{D}^r_t &= \text{Set of riders to be moved at time } t\text{, where } d \in \mathcal{D}^r_t \text{ corresponds to a trip with attribute } b_d, \\
\delta^b_{tdb} &= 1 \text{ if } d \in \mathcal{D}^r_t \text{ corresponds to a trip with attribute } b\text{, 0 otherwise,} \\
\mathcal{D}^e &= \text{Set of locations, where } d \in \mathcal{D}^e \text{ refers to a decision to move to a location in } \mathcal{D}^e, \\
\mathcal{D}_t &= \mathcal{D}^s \cup \mathcal{D}^r_t \cup \mathcal{D}^e.
\end{aligned}$$

We then capture decisions with

$$\begin{aligned}
x_{tad} &= \text{Number of times we act on a resource with attribute } a \text{ with a decision } d \in \mathcal{D}_t, \\
x_t &= (x_{tad})_{a \in \mathcal{A}, d \in \mathcal{D}_t}.
\end{aligned}$$

Decisions will be made according to a policy $X^\pi(S_t)$ which we design below. The policy has to return a vector x_t that satisfies the constraints

$$\begin{aligned}
\sum_{d \in \mathcal{D}_t} x_{tad} &= R_{ta}, \quad a \in \mathcal{A} \\
\sum_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}^r_t} x_{tad} \delta^b_{tdb} &\leq D_{tb}, \quad b \in \mathcal{B}.
\end{aligned}$$

We let $\mathcal{X}_t$ represent the set x_t that satisfies these constraints.

We capture the served customers using

$$\mathcal{D}^c_t(x_t) = \{b \in \mathcal{D}^r_b, \text{where } x_{tad}\delta^b_{tdb} > 0\}.$$

We do not require that we cover all the orders, although we anticipate that in a properly calibrated system, a very high percentage would be covered.

Exogenous information:
The main source of exogenous information is customer orders, which we model using

$$\hat{D}^r_t = \text{Set of new orders that first becomes known by time } t.$$

We are going to initially model our exogenous process as completely independent of the system. However, we could extend our model to one where customers look at an app to check on the availability of cars.

We may also introduce other sources of randomness such as cars failing or travel delays due to weather. We let W_t be our general set of exogenous information variables, including $\hat{D}_t$, capturing all information that first becomes known (from an exogenous source) by time t. The sequence $W_1, W_2, \ldots, W_T$ represents our exogenous information process, where $\omega \in \Omega$ is a complete sample sequence.

Transition function:
We begin by defining

$$\begin{aligned}
\delta^R_{aa'}(d) &= 1 \text{ if decision } d \text{ acting on a resource with attribute } a \text{ produces a resource} \\
&\quad \text{with attribute } a'\text{, 0 otherwise.}
\end{aligned}$$

The resource transition function is given by

$$R_{t+1,a'} = \sum_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} x_{tad} \delta^R_{aa'}(d) \tag{1}$$

The updated set $\mathcal{D}^r_{t+1}$ is given by

$$\mathcal{D}^r_{t+1} \quad = \quad \left(\mathcal{D}^r_t \setminus \mathcal{D}^c_t(x_t)\right) \cup \hat{D}^r_{t+1}. \tag{2}$$

Equations (1) and (2) make up our transition function, which we write generally as

$$S_{t+1} = S^M(S_t, x_t, W_{t+1}).$$

Objective function:
We wish to maximize total revenues from served customers minus operating costs which we view as a linear function of total distance traveled (moving customers, moving empty, and moving to and from charging stations), which we represent using $C(S_t, x_t)$. The objective may also include penalties for unserved customers. Our objective function is then

$$\max_{\pi} \mathbb{E} \left\{ \sum_{t=0}^{T} C(S_t, X^\pi(S_t)) | S_0 \right\}. \tag{3}$$

where $S_{t+1} = S^M(S_t, x_t, W_{t+1})$. Now we have the challenge of identifying a policy.

2 Designing a dispatch policy

There are two fundamental strategies for designing policies for sequential decision problems:

1) Policy search - Here we search over a parameterized class of policies, which may come in one of two forms: parametric function approximations (analytical functions that map states to actions), and policies based on parameterized cost functions (and constraints) that can be written

$$X^{CFA}(S_t|\theta) = \arg \max_{x_t \in \mathcal{X}^\pi_t(\theta)} \bar{C}^\pi(S_t|\theta)$$

where $\bar{C}^\pi(S_t|\theta)$ represents a modified cost function (such as a myopic policy for assigning the closest customer, possibly with penalties to encourage specific behaviors), subject to a possibly modified set of constraints $\mathcal{X}^\pi_t(\theta)$ (for example, which might capture estimates of forecasted customers). CFAs are widely used in industry practice, and represent the class of policy currently used by companies such as Uber and Lyft. CFAs can include deterministic forecasts of customers that might call in.

2) Policies based on lookahead approximations - These are policies that are trying to optimize over the contributions now and in the future. The most general form of this class of policy is

$$X^*_t(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + \mathbb{E} \left\{ \max_{\pi} \mathbb{E} \left\{ \sum_{t'=t+1}^{T} C(S_{t'}, X^\pi_{t'}(S_{t'})) \middle| S_{t+1} \right\} \middle| S_t, x_t \right\} \right). \tag{4}$$

Equation (4) is almost never computable (a decision tree is an example where this is possible). There are two general strategies for solving this equation. One is to replace the lookahead model with a lookahead approximation; this is the approach taken in stochastic programming. The other is to replace the lookahead term with a value function approximation, giving us a policy of the form

$$X^{VFA}_t(S_t) = \arg\max_{x} \left(C(S_t, x) + \overline{V}_t(S^x_t) \right), \tag{5}$$

While the parametric CFA describes the approach used in industry practice, we are going to use the VFA-based policy in equation (5). We use a value function approximation that depends only on the post-decision resource state R^x_t, which captures the future destinations of vehicles after a decision is made, but ignores uncovered demands. Thus, we use $\overline{V}_t(S^x_t) = \overline{V}_t(R^x_t)$. The next challenge is to design the functional form of $\overline{V}_t(R^x_t)$.

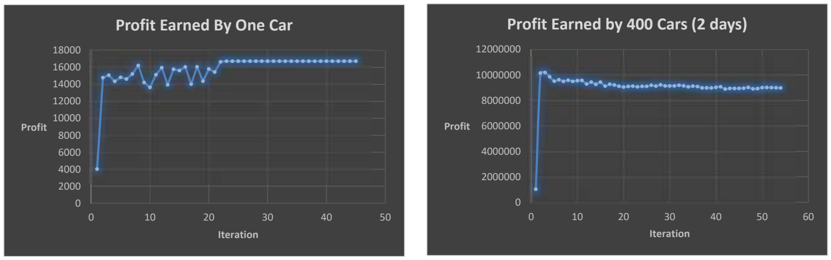


Figure 1: Performance of VFA policy after n iterations with 1 car (left) and 400 cars (right).

3 Experimental testing

To estimate $\overline{V}_t(R_t^x)$, we first assign vehicles at time t, and then let $\hat{v}_{ta}$ be the value of assigning a car with attribute a at time t (this can involve assigning the car to do nothing). The logic was tested using a sample of trips from a large database consisting of all trips made in New Jersey on a particular day. The test was first run on a fleet with one car, and then with 400 cars. After training for n iterations, we stopped and simulated the policy to assess its performance. The results are shown in figure 1.

With one car, the performance generally improved, but with 400 cars, there was rapid improvement followed by partial but steady degradation. We believe the reason for this behavior with larger fleets is that the value $\hat{v}_{ta}$ for a particular car was the average value, not the marginal value. Hence, this would work fine with one car (or very small fleets), but not larger fleets.

We have since replaced the original logic of assigning each car in a greedy fashion with logic which solves a global assignment of all cars, just as was used in Simao et al. (2009) and Bouzaiene-Ayari et al. (2014). This logic will do a better job of allocating cars spatially. For example, we may hold a car (or move it toward a zone with higher call-in density), rather than assign it to a customer, and instead use a car that may be farther away, but which has fewer opportunities.

This logic lets $\hat{v}_{ta}$ be the marginal value of a driver with attribute a, which is then used to estimate a value function approximation that captures the marginal value as a function of the amount of energy in storage. The marginal value $\hat{v}_{ta}$ can be computed using dual variables or numerical derivatives (which is more reliable but much more expensive). We will then use a layered value function approximation where we estimate a nonlinear VFA as a function of the number of drivers in a zone. This use of layered VFAs is similar to the strategy used in Bouzaiene-Ayari et al. (2014).

We will report on the results of our validation of the logic for controlling the fleet through value function approximations. We will compare different types of value function approximations, where the challenge is capturing the marginal value of a car with a certain charge level in a zone with a mixture of cars, each with their own charge level. We will then use the VFA that produces the best results to test different distributions of battery sizes across the fleet. Then, we hope to also demonstrate our ability to design the best possible distribution of battery sizes using the value functions as a guide.

References

Bouzaiene-Ayari, B., Cheng, C., Das, S., Fiorillo, R. & Powell, W. B. (2014), 'From Single Commodity to Multiattribute Models for Locomotive Optimization : A Comparison of Optimal Integer Programming and Approximate Dynamic Programming', *Transportation Science* pp. 1–24.

Cheung, R. K.-M. & Powell, W. B. (1996), 'An Algorithm for Multistage Dynamic Networks with Random Arc Capacities, with an Application to Dynamic Fleet Management', *Operations Research* **44**, 951–963.

Crainic, T. G. (2003), Long-Haul Freight Transportation, *in* R. Hall, ed., 'Handbook of Transportation Science', Springer, pp. 451–461.

Jordan, W. C. & Turnquist, M. A. (1983), 'A Stochastic Dynamic Network Model for Railroad Car Distribution', *Transportation Science* **17**, 123–145.

Powell, W. B. (1987), 'An operational planning model for the dynamic vehicle allocation problem with uncertain demands'', *Transportation Research 21B* pp. 217–23253.

Powell, W. B. & Godfrey, G. A. (2002), 'An adaptive dynamic programming algorithm for dynamic fleet management, I: Single period travel times', *Transportation Science* **36**(1), 40–54.
URL: *http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:An+Adaptive+Dynamic+Programming+Alg*

Powell, W. B., Sheffi, Y., Nickerson, K. S., Butterbaugh, K. & Atherton, S. (1988), 'Maximizing Profits for North American Van Lines' Truckload Division: A New Framework for Pricing and Operations', *Interfaces* **18**, 21–41.

Powell, W. B. & Topaloglu, H. (2005), Fleet Management, *in* S. Wallace & W. Ziemba, eds, 'Applications of Stochastic Programming', Math Programming Society - SIAM Series in Optimization, Philadelphia, pp. 185–216.

Simao, H. P., Day, J., George, A. P., Gifford, T., Powell, W. B. & Nienow, J. (2009), 'An Approximate Dynamic Programming Algorithm for Large-Scale Fleet Management: A Case Application', *Transportation Science* **43**(2), 178–197.

Topaloglu, H. & Powell, W. B. (2006), 'Dynamic-Programming Approximations for Stochastic Time-Staged Integer Multicommodity-Flow Problems', *INFORMS Journal on Computing* **18**(1), 31–42.

From 'No Data' to 'Some Data' to 'Big Data' Towards a Cyber-Physical System for Proactive Traffic Management

Pitu Mirchandani
School of Computing, Informatics & Decision Systems Engineering, Arizona State University
Email: pitu@asu.edu

Kerem Demirtas
School of Computing, Informatics & Decision Systems Engineering, Arizona State University
Email: Kerem.Demirtas@asu.edu

Viswanath Potluri
School of Computing, Informatics & Decision Systems Engineering, Arizona State University
Email: vpotlur1@asu.edu

Extended Abstract

The objective of the research reported is to synergistically use a cyber-physical infrastructure consisting of smart-phone devices; cloud computing, wireless communication, and intelligent transportation systems to manage diverse vehicles in the complex urban network – through the use of traffic controls, and route guidance to jointly optimize drivers' mobility and the sustainability goals of reducing energy usage and improving air quality. The system being developed, ***MIDAS,*** *is to* proactively **m**anage the **i**nteracting traffic **d**emand **a**nd the available transportation **s**upply. A key element of ***MIDAS*** is the data collection and display device ***PICT*** that collects each participating driver's vehicle **p**osition, forward **i**mages from the vehicle's dashboard, and **c**ommunication **t**ime stamps, predict traffic ahead and provide signal controls and route advisories to optimize given objectives.

Given the increasing congestion in most of the urban areas, and the rising costs of constructing traffic control facilities and implementing highway hardware, ***MIDAS*** could revolutionize the way traffic is managed on the urban network since all computing is done via clouds and the drivers instantly get in-vehicle advisories and predicted conditions. This multidisciplinary project is at the cutting edge in several areas: real-time image processing, real-time traffic prediction and supply/demand management, and cloud computing. Historically, setting signals and ramp-metering rates at time-of-day timings, used **no** on-line data; actuated and semi actuated signals with usual detectors (loop detectors and /or video detectors) allowed us to use **some** on-line data. ***MIDAS*** takes us from some date to "**big data**" in that it fully exploits the streaming data available from smart phones and other sensors, remote and on-the-infrastructure sensors.

Based on the information from the PICT devices, we obtain Lagrangian measurements (explained later) that allows us to predict traffic on **each lane** and subsequently estimate the travel time on each link and lane, allowing the vehicles to select lanes and routes to reach their destinations. Subsequently, based on these routes, the ***MIDAS*** traffic control logic proactively sets signal phases to minimize delays and/or queues at the intersections. These delays in turn can provide better route guidance. The iteration between phase durations at the intersections and route predictions/guidance simultaneously sets these decisions to optimize the jurisdictions traffic objectives.

Traffic prediction

As in most traffic management applications, real-time data collection plays a crucial role for a reliable and robust traffic management system, since it provides a means to partially monitor the facilities of the network. Inductive loop detectors, which are installed under the pavement at certain locations of the transportation facilities, have been the major source of data for traffic management applications since the 1960s [KFM90]. This type of data is often referred to as *Eulerian*, since the data is obtained from Eulerian sensors that are fixed at points in space. On the other hand, mobile sensors that move with the flow of traffic provide another type of data often referred to as *Lagrangian* data, which is structurally quite different than Eulerian data.

In this research, we use the Lagrangian measurements in a more effective way based on techniques and the ideas from Daganzo's [Dag06] and Newell's models [New02, New93a, New93b, New93c]. applied in a Lagrangian framework. Moreover, to the best of the authors' knowledge, there is no study yet, that investigates the traffic state estimation and prediction problem at a **lane level resolution**, which we believe to be important, especially relevant to new technological developments in intelligent vehicles such as connected and autonomous vehicles. Multiple traffic flow models that consistently apply to both microscopic and mesoscopic levels of traffic will be combined to have achieve high fidelity predictions at all resolutions.

Traffic prediction and estimation algorithms are being developed that use PICT data – such measurements are Lagrangian since vehicles within PICT images provide information on trajectories. For quick traffic state prediction, mostly for dynamic travel times, a novel prediction algorithm is being investigated. The new algorithm takes input from existing online state estimation methods, and produce both macroscopic traffic state variables of interest and reliable individual travel times that are dynamically updated in flexible time horizons. A Kalman filter framework has been developed at the microscopic level resolution. The proposed algorithm can identify both interdriver and also intradriver heterogeneity in terms of microscopic driver behavior parameters, which collectively, have a significant effect on mesoscopic and even macroscopic conditions on the whole network. As opposed to aggregating the study to the links, the new estimation and prediction algorithm focus on **individual lanes** of the links, which makes the research contribution novel as the current practices and theory are mostly limited to the link level.

Traffic control

In any proactive traffic control system, the signal control algorithm is the key component in determining the optimal phase sequence and phase durations, by minimizing some user defined traffic performance measure, such as minimizing total delays, stops or queues at the intersections. Many researchers have worked towards developing an adaptive traffic control system in the past, of which ***RHODES*** [Mir01] is a better performing real time adaptive traffic control system and also has been field tested. But like any other existing adaptive traffic control systems, ***RHODES*** uses data from fixed loop detectors for prediction and estimation. On the other hand, ***MIDAS*** traffic control installed at each intersection uses Lagrangian data from PICT devices from all the approaches, combined with traffic estimation methods (as discussed above), to predict approaching vehicles' individual arrival times, platoon movements, lane based traffic, and to estimate turning ratios at the intersection. As these variables get updated dynamically for a user defined time horizon, ***MIDAS*** traffic control algorithm feeds on the updated predictions to estimate the queues in each lane on all approaches of each intersection, at each decision epoch with corresponding time horizons.

MIDAS signal control algorithm uses forward recursion Dynamic Programming (DP), to minimize the user defined performance measure over a finite-time horizon that rolls forward and, then uses a backward recursion to retrieve the optimal phase schedule. The signal control algorithm of ***MIDAS*** is similar to the ones in [Sen97] and [Mir01] but with an efficient data structure. **MIDAS** signal control algorithm determines the optimal signal phase sequence and duration of each phase in the sequence, by taking user defined set of phases (any number of phases) and time horizon as input parameters. Control algorithm runs DP at some time stamp 't', with prescribed time horizon 'T' and, considering the tentatively estimated arrivals on all approaches of the intersection over the timeline 't+T'. The DP is formulated such that each "stage" of the DP is associated with a signal phase and number of time units allocated to all past phases before the current stage is defined as stage variable. The DP solution consists of the phase sequence and time units allocated to each phase over the next 'T' time units. Fig.1 gives an illustration. At every DP run, the sequence of phases begins with the current phase that is green at the intersection, which allow for the phase to be terminated or extended based on the updated observations.

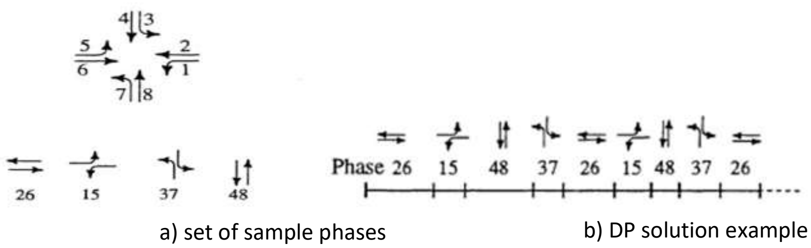


Figure 1: *Illustration of movements, phases and a DP solution on for a sequence of phases*

The underlying data structures of the algorithm are designed to hold tentative queues (vehicles waiting at upstream intersections) and committed arrivals (on their way to the intersection) separately and also capture individual vehicular information like waiting times, number of stops and other parameters so that the DP is able to minimize any prescribed traffic performance measure. ***MIDAS*** traffic control algorithm is designed to optimize any given objective for each particular intersection, and honors the constraints of each intersection like minimum green time, maximum green time, phase sequence restrictions, etc.

Joint optimization of traffic control and vehicle routes

Most traffic signal optimization (TSO) methods, whether offline or online, such as described above, assume link flows (or measure them in online methods) and assume turning ratios, also referred to as turning proportions, of the arriving flows at the intersections. Implicit traffic assumptions in most of the underlying optimization models is that the flows for the signal timing planning horizon are stationary and, thus, signal timings are determined for the **given** flows and turning proportions. In the combined Route Guidance and Traffic Signal Optimization (RGTSO) model described in this paper, the phase status of each traffic signal is modeled with variables describing the movement allowed at the scheduled phase. Link travel time times depend on the number of vehicles routed through each link as per traffic flow theory. Each vehicle traveling from its origin to its destination travels each link in a candidate route with travel time depending upon the resultant vehicle flow on the link; the waiting, if any, at the intersection, explicitly depends on the queues and the signal status. The space-time trajectory of each vehicle through the network will give its total travel time and the optimization objective value becomes the total travel times of all guided vehicles which can be iteratively decreased by changing guided routes and/or signal phase schedules. Other optimization criteria could be considered such as minimizing number of stops to improve coordination, minimizing average queue sizes, minimizing fuel consumption, etc. Decomposition of the overall optimization problem results in two subproblems: TSO and route guidance (RG); overall optimization solution approach iterates between TSO and RG problems. Some previous attempts for solving jointly TSO and RG problems within a long-term static traffic equilibrium setting have been reported by Chen and Ben-Akiva [Che98], Smith and Mounce [Smi11], Aziz and Ukkusuri [Azi12]. This research is among the first two do so for applicability for real-time optimization with time-resolution of seconds. The output of the RGTSO is second-by second guidance for each vehicle on system optimal routes and concurrent second-by-second phase setting of traffic signals in the network. Figure 2 illustrates the underlying coupled space-time and phase-time networks for a 9-intersection network. Some equilibrium analysis of the solution of RGSTO for this network are reported by. Li et al. [Li15].

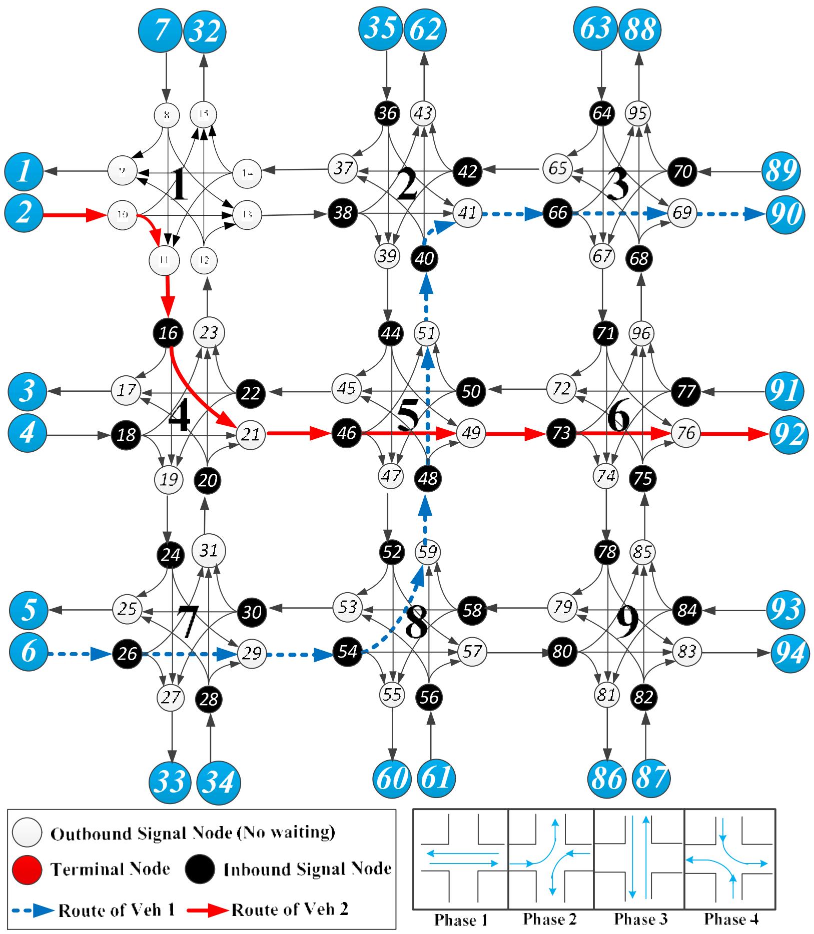


Figure 2. *Space-time-phase trajectories illustration*.

For illustrative purposes, two vehicles space-phase trajectories are shown in the figure, solid (red) route and dashed (blue route). If the time dimension was included it could be that these vehicles arrive at Intersection 5 at the same time. Resolution of the conflict could be done either through changing phase times at the intersection and/or changing the routes to their respective destination. Solution of RGTSO would provide the optimum conflict-free decisions**.**

Concluding Remarks

The research team is currently working on the above described MIDAS system. It is anticipated that by the time of the TSL conference, more details and analysis will be reported and some additional results on a simulation-based platform with more general real networks will be reported

References

[Azi12]. Aziz, H.M.A., Ukkusuri, S.V., 2012. “Unified framework for dynamic traffic assignment and signal control with cell transmission model”. *Transportation Research Record 2311*, 73-84.

[Che98] Chen, O., Ben-Akiva, M., 1998. “Game-Theoretic Formulations of Interaction between Dynamic Traffic Control and Dynamic Traffic Assignment”. Transportation Research Record 1617, 179-188.

[Dag06] Daganzo, Carlos F. "In traffic flow, cellular automata= kinematic waves" *Transportation Research Part B*: *Methodological* 40.5 (2006): 396-403.

[KFM90] James H Kell, Iris J Fullerton, and Milton K Mills. Traffic detector handbook. Technical Report, 1990

[Li15] P Li, P Mirchandani, X Zhou, “Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetworks”, *Transportation Research Part B: Methodological*, 81, 103-130, 2015

[Mir01] Mirchandani, P., and Head, L. “A real-time traffic signal control system: Architecture, algorithms and analysis.” *Transp. Res., Part C: Emerging Technologies*, 9, (2001). 415–432.

[New02] Newell, Gordon F. "A simplified car-following theory: a lower order model." *Transportation Research Part B: Methodological* 36.3 (2002): 195-205.

[New93a] Newell, Gordon F. "A simplified theory of kinematic waves in highway traffic, part I: General theory." *Transportation Research Part B: Methodological* 27.4 (1993): 281-287.

[New93b] Newell, Gordon F. "A simplified theory of kinematic waves in highway traffic, Part II: Queueing at freeway bottlenecks." *Transportation Research Part B: Methodological* 27.4 (1993): 289-303.

[New93c] Newell, Gordon F. "A simplified theory of kinematic waves in highway traffic, part III: Multi-destination flows." *Transportation Research Part B: Methodological* 27.4 (1993): 305-313.

[Sen97] Sen, U., and Head, K. L. “Controlled optimization of phases at an intersection.” *Transportation Science.* 31(1), (1997). 5–17.

[Smi11] Smith, M., Mounce, R., 2011. “A splitting rate model of traffic re-routeing and traffic control”. *Transportation Research Part B: Methodological* 45, 1389-1409.

Multimodal Transportation Services
TA1: Air Traffic and Air Networks
Thursday 8:30 – 10:30 AM
Session Chair: David Lovell

8:30 Choice-Based Airline Fleet Assignment
[1]Chiwei Yan, [1]Cynthia Barnhart, [2]Vikrant Vaze*
[1]Massachusetts Institute of Technology, [2]Dartmouth College

9:00 The Most Reliable Path Problem for Airline Travel with Connections
[1]Michael Redmond, [1]Ann Campbell, [2]Jan Ehmke*
[1]University of Iowa, [2]Europe University Viadrina

9:30 Modeling Flight Delay Propagation: A New Analytical-Econometric Approach[†]
*Nabin Kafle, Bo Zou**
University of Illinois at Chicago
[†]No extended abstract included

10:00 Sources of Flight Inefficiency in the National Airspace System: An Econometric Approach
[1]Mark Hansen, [1]Yulin Liu, [2]Michael Ball, [2]David Lovell, [2]Cara Chuang*
[1]University of California-Berkeley, [2]University of Maryland

Choice-Based Airline Fleet Assignment

Chiwei Yan[1], Cynthia Barnhart[1], and Vikrant Vaze[2]

[1]Operations Research Center, MIT
[2]Thayer School of Engineering, Dartmouth College

1 Introduction

Assigning aircraft types to the flight legs in an airline's schedule is an important tactical decision which greatly impacts airline's profit (Barnhart et al. 2002). Existing fleet assignment approaches effectively capture fleet availability constraints, aircraft maintainence constraints, and network flow balance constraints. However, customer demand is usually modelled in a simplified way in these existing approaches. One commonly made assumption, called the independent demand assumption, states that each passenger has a unique itinerary product that he/she intends to buy, and if that product is not available due to capacity constraints or revenue management policies, then the demand is simply lost. Such an assumption is not valid in practice because there always exist substitution effects among similar itinerary products. A passenger who is not able to buy his/her favorite itinerary product might choose an alternative product instead.

Motivated by this fact, in this research, we study a new fleet assignment model (FAM) where customer demand interactions are captured using discrete choice models. Discrete choice models are commonly used in marketing literature to model product substitutions (McFadden 1980) and in transportation literature to model travel demand (Ben-Akiva and Lerman 1985). Recently, they are also being widely incorporated into airline revenue management studies (Talluri and van Ryzin 2004, Liu and van Ryzin 2008, Gallego et al. 2014). However, there is little work in incorporating choice models into airline planning models. Wang et al. (2014) was one of the first research studies where multinomial logit (MNL) choice is incorporated into fleet assignment and schedule design. As revealed in the paper, the downside of a straightforward combination of FAM with MNL choice is loss of tractability because of the dramatic change to the structure of the FAM model. This issue is further exacerbated with other advanced choice models. From our own experience, for a problem instance from a major US airline, the straightforward model directly combining FAM and MNL choice does not produce even a feasible solution in 30 hours of computation time with a state-of-the-art commercial solver. This computational burden is the major obstacle which prevents choice-based FAM (CFAM) from being applied in the airline industry. Faced by this difficulty, our research makes the following contributions to CFAM:

1. We develop a reformulation, decomposition and approximation scheme for CFAM. The approximation scheme has the capability to specify the degree of balance between computational efficiency and solution quality.

2. The developed approach naturally separates the revenue calculation and fleet assignment into a two-step process so that varying the choice model assumptions will not change the structure of the fleet assignment problem. This enables our approach to incorporate more advanced choice models, such as mixed MNL, ranking-based model (Farias et al. 2013), etc., without a dramatic increase in computational time.

In summary, the proposed approach provides a practically efficient framework to incorporate customer choice into fleet assignment models. With more accurate customer demand modeling, higher profit is expected under this approach. This is confirmed by our preliminary computational experiments showing superior profits as well as faster computational performance. It should also be noted that the proposed framework can be extended to other airline planning models where customer demand plays a key role, such as the schedule design problems.

2 Methodology

We utilize an existing reformulation called subnetwork-based FAM (Barnhart et al. 2009) to address choice-based FAM. The subnetwork-based FAM (SFAM) is an approximation scheme originally developed for solving itinerary-based FAM (Barnhart et al. 2002) efficiently. The key idea of SFAM is to utilize composite variables to model fleet assignment decisions. In a standard FAM, binary variables $x_{l,f}$ are defined to equal 1 if fleet type f is assigned to flight l, and 0 otherwise. In SFAM, flights are first partitioned into different subnetworks. For each subnetwork k, we enumerate all possible fleet assignments for all the flights in it. We then use a binary variable w_j^k which equals 1 if fleet assignment j is chosen for subnetwork k, and 0 otherwise. The following table shows an example of a subnetwork consisting of two flight legs (l_1 and l_2) and all possible fleet assignments with two fleet types (A and B). As can be seen from the table, there are four possible assignments for this subnetwork, and the assignment variable w_j^k here indicates whether a particular one is adopted.

Flight	Assign. 1	Assign. 2	Assign. 3	Assign. 4
l_1	A	A	B	B
l_2	A	B	A	B

Table 1: Illustration of fleet assignment of a subnetwork

With this new definition of the fleet assignment variables, SFAM represents a change-of-variable approach relative to the standard FAM formulation. As pointed out in Barnhart et al. (2009), this Dantzig-Wolfe like reformulation in SFAM enjoys better computational efficiency because of the tightened LP relaxation bound. The advantage of using SFAM under passenger choice behavior models is that it can naturally separate the fleet assignment and revenue calculations into a two-step process where the revenue associated with each possible fleet assignment can be calculated offline so that the structure of FAM is independent of the assumptions made by the choice models.

On the other hand, the key challenge of SFAM is that the required number of assignment variables grows exponentially with the size of the subnetwork. Thus the size of the subnetwork determines the key trade-off between computational efficiency and solution quality, where coarser partition and larger subnetwork size will lead to greater solution quality but higher computational requirements, and vice versa. Barnhart et al. (2009) developed heuristic approaches to find good subnetwork partitions. After the subnetwork is fixed, if there is a fare product shared by multiple subnetworks, Barnhart et al. (2009) divide the price of that fare product across subnetworks with some relatively non-sophisticated methods. However, since the original SFAM assumes independent demand, the approximation scheme in Barnhart et al. (2009) is not tight enough for CFAM. Thus, we develop a novel linear program for splitting the fare of these cross-subnetwork products to find an optimal fare structure for each subnetwork to further enhance its approximation quality.

3 Computational Experiments

We conduct preliminary computational experiments using two airline networks. The descriptive information on these two networks is provided in Table 1:

Network	Number of Flights	Number of Fare Products	Schedule Repeatition
1	184	10000	Daily
2	1400	100000	Weekly

Table 2: Two testing networks

For each network, we assume there is a ground-truth MNL model which governs the customer demand. We then solve CFAM (Wang et al. 2014), IFAM (Barnhart et al. 2002) and our proposed choice-based SFAM for these two networks. For IFAM, we use the first-choice demand from the MNL choice model as the unconstrained demand for each product. For each of the fleeting solutions coming out of these models, we run a choice-based linear program for network revenue management under the ground-truth MNL choice model (Gallego et al. 2014) to evaluate the profit. We summarize the test results in Tables 2 and 3.

In summary, we find that for the small network instance (network 1), the proposed choice-based SFAM approach is able to produce near-optimal solutions with significantly shorter computation time; and for the large network

Model	Daily Profit	Annual profit improvement over IFAM	CPU Time
IFAM	\$922,400	-	60 sec
CFAM	\$936,800	5.25 million	7200 sec
SFAM	\$936,790	5.25 million	900 sec

Table 3: Testing results of network 1

Model	Weekly Profit	Annual profit improvement over IFAM	CPU Time
IFAM	\$10,816,837	-	5 hr
CFAM	No Solution	-	30 hr
SFAM	\$10,850,641	1.76 million	5 hr

Table 4: Testing results of network 2

instance (network 2), it is able to produce good solutions within reasonable computational budget.

4 On-going Work

We are currently extending and testing the choice-based SFAM approach under advanced choice models including mixed multinomial logit and ranking-based choice models. Our ultimate goal is to evaluate this approach under a data-driven setting: given customer shopping and transaction data, we first estimate various types of choice models; we then benchmark the fleeting solution from our SFAM approach using different choice models to quantify: 1) how much profit we can gain by using a more advanced and accurate demand model; 2) what is the associated increase in computational requirement; and 3) how much better our profits are compared with less sophisticated approaches. This holistic approach will move forward the state-of-the-art for airline fleet assignment under customer choice behaviors and provide a practical planning tool for airlines to build effective fleeting decisions directly from transaction and shopping data.

References

Barnhart, C., Farahat, A., and Lohatepanont, M. (2009). Airline fleet assignment with enhanced revenue modeling. *Operations Research*, 57(1):231–244.

Barnhart, C., Kniker, T. S., and Lohatepanont, M. (2002). Itinerary-based airline fleet assignment. *Transportation Science*, 36(2):199–217.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT Press.

Farias, V. F., Jagabathula, S., and Shah, D. (2013). A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322.

Gallego, G., Ratliff, R., and Shebalov, S. (2014). A general attraction model and sales-based linear program for network revenue management under customer choice. *Operations Research*, 63(1):212–232.

Liu, Q. and van Ryzin, G. (2008). On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2):288–310.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, pages S13–S29.

Talluri, K. and van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33.

Wang, D., Klabjan, D., and Shebalov, S. (2014). Attractiveness-based airline network models with embedded spill and recapture. *Journal of Airline and Airport Management*, 4(1):1–25.

The Most Reliable Path Problem for Airline Travel with Connections

Michael Redmond, Ann Melissa Campbell
Department of Management Sciences, University of Iowa
Jan Fabian Ehmke
Business Analytics Group, Europe University Viadrina

1 Introduction

Travel between many cities can require multiple legs, such as several trains, bus rides, or flights, to arrive at the final destination. When there are many different combinations available to travel from origin to destination (journeys), travelers are often presented with different options that minimize price, total travel time, or a combination of these two. In choosing between options for such multiple leg journeys, customers would like to avoid a journey that will likely lead to missed connections or arriving late to their destination. As Börjesson and Eliasson (2011) point out, simply knowing the average delay or the on-time arrival probability for individual legs will not give a complete picture of a journey's reliability. However, the reliability of the journey is not often available when making this a priori decision.

Since variability in travel times can create delays and missed connections that greatly affect travelers' probability of arriving on-time to their destinations, identifying and presenting reliable paths is also important. Identifying the most reliable path utilizing travel time data is not simple, and new techniques are required to do this efficiently. We propose an approach to evaluate the reliability of a journey from origin to destination given a start time and travel time budget. Specifically, we focus on airline travel and making a priori evaluations of reliability for different journeys. Airline travel is unique in that the network inherently exhibits a high amount of uncertainty with over 20% of flights experiencing delays or cancellations in 2015 (USDoT, 2016). We present our model of reliability in Section 2, explore publicly available flight data in Section 3, and discuss techniques of finding the most reliable path in Section 4.

2 Reliability Model

In this section, we define the reliability of a multi-leg journey and discuss how to compute the reliability measure.

2.1 Problem Definition

We aim to identify the reliability of different multi-leg journeys from an origin to a destination. We define reliability based on

- λ^{od}: specific journey from origin to destination where λ represents a sequence of legs that make up that journey,
- $start\_time$: the time the passenger is able to begin his or her journey, and
- B: the travel time budget for reaching the destination. The latest arrival time at the destination will be $start\_time + B$.

The reliability of a particular sequence λ^{od} will be expressed by $rel(\lambda^{od}, start\_time, B)$ and represents the probability that this sequence will lead to an arrival at the destination by $start\_time + B$. For simplicity of presentation, we will define the sequence λ as {leg 1, leg 2, ... leg m}, where m is the number of legs in the given journey. For each leg i in a journey, we assume we know the following:

- $sched\_d_i$: the time leg i is scheduled to depart
- min_dur_i: the minimum travel time of leg i
- $sched\_a_i$: the time leg i is scheduled to arrive at its destination.

Connecting two legs requires a feasible transfer. Feasibility of connections between legs is tied to a constant value min_c, the minimum time necessary to transfer from one leg to the next, such as the time to walk between terminals in an airport. We assume that if an arrival allows min_c time to occur before the next departure, the connection will be made. We assume each sequence under consideration (λ) is feasible with regard to yielding a path from an origin to the destination and planned transfer times between legs (e.g. for leg 2, it must be true that $sched\_a_1 + min\_c \leq sched\_d_2$).

We define the most reliable path problem as the problem to identify the sequence λ that maximizes $rel(\lambda^{od}, start\_time, B)$ for a given $start\_time$ and budget B.

2.2 Computation of Reliability

Computing the reliability measure is challenging, since the reliability of the multi-leg journey depends on the probability of all connections being made as well as the arrival time of the last leg in the sequence. In our model, we will assume that the travel time of each leg is independent. We also assume that legs do not wait for late arriving passengers, and no leg i departs before $sched\_d_i$. The probability information for each leg i can be expressed by

- $P(D_i = t)$: probability that departure of leg i occurs at time t,
- $P(A_i = t)$: probability that arrival to end of leg i occurs at time t.

We can precompute these values using operational flight data from the United States Bureau of Transportation Statistics' (USBTS) historical database at *http://www.transtats.bts.gov*. Next, we will discuss how we use this data to compute $rel(\lambda^{od}, start\_time, B)$, for example. We start with the simplest case of one leg and show how to extend this to consider multi-leg journeys with connections.

2.2.1 One Leg

A sequence with only one leg (e.g. $\lambda^{od} = \{leg1\}$) makes it to its destination by time $start\_time + B$ with the following probability:

$$P(A_i \leq start\_time + B) = \sum_{t=sched\_d_i+min\_dur_i}^{start\_time+B} P(A_i = t) \tag{1}$$

Thus, $rel(\lambda^{od}, start\_time, B) = P(A_i \leq start\_time + B)$.

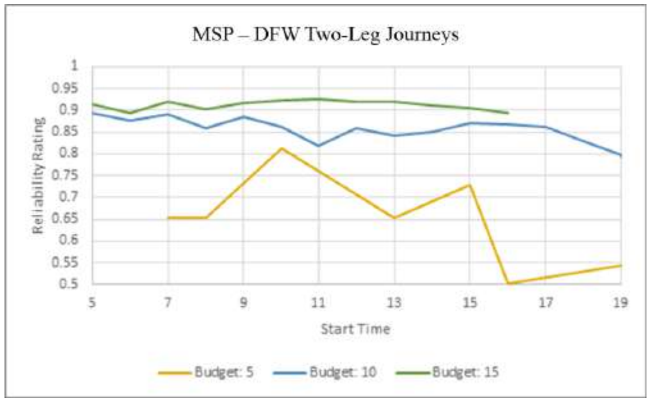
2.2.2 Two or More Legs

With two or more legs, we must compute the likelihood that each connection in the sequence is made as well as the arrival to the final destination occurs by $start\_time + B$. We assume a passenger cannot arrive to the destination by $start\_time + B$ unless they make all of the connections. In terms of probability, $PC_{1,2} = P(C_{1,2})$ will represent the probability the first connection is made, and $PC_{2,3} = P(C_{2,3} \mid C_{1,2})$ is the probability the second connection is made given the first connection is made, etc. Last, $P(A_m \leq start\_time + B \mid C_{1,2} \cap \ldots C_{m-1,m})$ will represent the probability that the arrival to the destination occurs within the travel time budget given all of the prior connections are made. Then, our final calculation of reliability becomes

$$\begin{aligned} rel(\lambda^{o,d}, start\_time, B) = P(A_m \leq start\_time + B \mid C_{1,2} \cap \ldots C_{m-1,m}) \\ \times \prod_{i,j} PC_{i,j} : i = 1, ..m-1, j = i+1. \end{aligned} \tag{2}$$

3 Data Exploration

Based on historical flight data, we explore the evolution of reliability measures for varying start times and travel time budgets for various origins and destinations. Figure 1 represents an example of how varying time budgets impact the reliability for the origin-destination pair *Minneapolis – Dallas Fort Worth* for different start times. The figure shows the average reliability rating of journeys departing in a particular hour of day. Note that it is based on information only from journeys that require at least one transfer. We can see that larger time budgets allow for better reliability ratings, since larger time budgets allow for longer layovers. Long layovers can offset delays that may interrupt connections.



4 Solution Approach

We are investigating several approaches that consider the above reliability measure in the computation of the most reliable path. We begin the network search by creating a simplified flight network containing a restricted set of arcs based on the value of *start_time*, adapting a concept from Delling et al. (2009) that helps us to keep the size of the network small (Phase I). In Phase I, a path begins at the terminal node of the origin airport A_o and proceeds to the departure node of Airport A_o. From there, all possible arcs from the departure node of A_o to each feasible airport arrival node are limited to the next available flight if a direct connection exists. The earlier and later flights between airports will be ignored. With the simplified flight network, we can use a variant of Dijkstra's algorithm to find a feasible path and compute the reliability of that shortest path, namely the initial incumbent most reliable path (MRP). Since Phase I does not consider the entire flight network, more reliable journeys may be overlooked.

In Phase II, we are expanding the network search to include all scheduled flights that occur throughout a day. Unfortunately, this expansion results in an exponential number of potential journeys between a particular origin and destination. Improvements we have been investigating:

- A *travel time based lower bound* in order to prevent the algorithm from considering nodes far from the destination.
- A *reliability lower bound* to prevent adding legs that will cause the reliability of a partial journey to be less than the reliability of an already discovered complete journey. Partial journeys are not added to the priority queue if they are not as reliable as this previously found complete journey.
- Implementing a k-shortest path algorithm, the highest reliability among these k shortest paths can be used as a lower bound on our reliability measure.

The acceleration techniques and k-shortest path algorithms allow us to expedite the network search. We will continue to improve the algorithm and metric to provide the traveler a sense of reliability in a network of uncertainty.

The following table highlights exemplary origin-destination pairs that both the Phase I and Phase II algorithms were able to navigate in order to find the most reliable paths. In addition to the start time and travel budget, the runtime of the algorithm is also displayed. *Max Rel I* represents the highest reliability found during Phase I, and *Max Rel II* likewise represents the reliability measure for the MRP in the expanded network.

Origin	Dest	Start time	Budget	MaxRel I	Time	MaxRel II	Time
Cedar Rapids	Nashville						
CID	BNA	5:00 AM	6 hours	89.9%	1.3 s	89.9%	1.0 s
CID	BNA	5:00 AM	8 hours	91.3%	10.9 s	95.3%	10.3 s
CID	BNA	5:00 AM	10 hours	93.6%	30.1 s	96.2%	30.0 s
CID	BNA	1:00 PM	6 hours	86.4%	0.9 s	86.4%	1.7 s
CID	BNA	1:00 PM	8 hours	90.2%	1.8 s	90.6%	7.6 s
CID	BNA	1:00 PM	10 hours	96.8%	6.1 s	96.8%	19.3 s
Friedman,ID	Miami						
SUN	MIA	5:00 AM	10 hours	66.4%	1.1 s	66.4%	4.0 s
SUN	MIA	5:00 AM	13 hours	78.6%	26.5 s	79.8%	77.0 s
SUN	MIA	5:00 AM	16 hours	83.0%	76.0 s	83.8%	201.0 s
Des Moines	San Diego						
DSM	SAN	5:00 AM	9 hours	93.1%	2.5 s	93.1%	4.7 s
DSM	SAN	8:00 AM	9 hours	76.8%	4.0 s	95.2%	8.7 s
DSM	SAN	11:00 AM	9 hours	96.9%	2.7 s	98.2%	6.9 s
DSM	SAN	2:00 PM	9 hours	97.1%	1.3 s	97.1%	3.7 s
Bangor, ME	Hilo, HI						
BGR	ITO	5:00 AM	24 hours	88.2%	66.0 s	N/A	N/A

On the trip from Cedar Rapids to Nashville, we can see the trends that we witnessed in the data exploration regarding the impact of increasing the travel budget. Likewise, this decrease in the reliability of the most reliable path is seen with small travel budgets. The 8:00 AM start time from Des Moines to San Diego demonstrates the value that expanding the network can have on finding the most reliable path. With the restricted network, the most reliable path was only able to arrive within the budget 77% of the time, but the expanded network found a journey that was 95% reliable. Finally, the long journey from Maine to Hawaii shows that the Phase I algorithm is able to find a reliable journey of 88%, but the Phase II algorithm requires more than the time allowed. This is the springboard for our future research.

References

M. Börjesson and J. Eliasson, "On the use of average delay as a measure of train reliability," *Transportation Research Part A: Policy and Practice* 45, 171–184 (2011).

D. Delling, T. Pajor, D. Wagner, and C. Zaroliagis, "Efficient Route Planning in Flight Networks," *ATMOS '09* 12 (2009).

USDoT, "Air Travel Consumer Report: February 2016," (2016).

SOURCES OF EN ROUTE FLIGHT INEFFICIENCY IN THE NATIONAL AIRSPACE SYSTEM: AN ECONOMETRIC APPROACH

Mark Hansen
National Center of Excellence for Aviation Operations Research (NEXTOR II)
114 McLaughlin Hall
University of California, Berkeley, CA 94720
Tel: 510-642-2880; Fax: 510-643-3955; Email: mhansen@ce.berkeley.edu

Yulin Liu
National Center of Excellence for Aviation Operations Research (NEXTOR II)
107D McLaughlin Hall
University of California, Berkeley, CA 94720
Tel: 510-590-7867; Email: liuyulin101@berkeley.edu

Cara Chuang
Department of Civil and Environmental Engineering
University of Maryland, College Park, MD 20742
Email: cara.y.chuang@gmail.com

David Lovell
Department of Civil and Environmental Engineering
1173 Glenn L. Martin Hall
University of Maryland, College Park, MD 20742
Tel: 301-405-7995; Fax: 301-405-2585; Email: lovell@umd.edu

Michael O. Ball
Robert H. Smith School of Business
2416 Van Munching Hall
University of Maryland, College Park, MD 20742
Tel: 301-405-2227; Email: mball@rhsmith.umd.edu

ABSTRACT

We investigate causes of en route flight inefficiency for US domestic flights into and out of 34 major US airports, using a dataset of several million flights from the years 2013 and 2014. Following earlier work, our inefficiency metrics compare the distance flown between airport terminal exit and entry points with the achieved distance, and further isolate the effects of pre-specified, and often not ideal, entry and exit points (TMA) and excess distance flown between these points (DIR). We find the TMA inefficiency decreases with flight distance, while DIR inefficiency is roughly constant with distance. Inefficiency varies considerably for flights between a given airport-pair, with median values generally less than 5%. To assess causes of inefficiency, we employ two different methodologies, both of which are based on clustering individual trajectories and identifying nominal trajectories within each cluster. In the first approach, the nominal trajectories are used to compute aggregate metrics related to different causal factors, for example exposure to convective weather. These metrics are then used as explanators in regression models in which inefficiency metrics are the dependent variables. In the second approach, we model the cluster to which a given trajectory belongs based on characteristics of nominal routes and then estimate regression models which take into account the trajectory cluster as well as metrics related to the various causal factors. Results for inefficiency causes including convective weather, winds, and TMI actions are presented.

Keywords: Flight Performance, En route inefficiency, Fixed effects model

1. INTRODUCTION

Recent years have witnessed growing interest in comparative performance assessment of air navigation service providers. Such efforts face difficult challenges, in particular the need to identify key performance indicators (KPIs) that are precisely defined, can be assessed using data available to all providers, and which capture the major dimensions of aviation system performance. The potential payoffs from comparative performance assessment more than justify such work, however. These include identifying opportunities for performance improvement, determining the benefits from modernization, and more fundamentally understanding the linkages between structure and performance in the air navigation service domain.

Comparative assessments typically emphasize macro comparisons. For example, the most recent comparison of US and European air navigation system performance *(1)* emphasizes conclusions such as "Europe continues to demonstrate less additional time in the taxi-out phase than in the US" and "the US continues to show a lower level of inefficiency in the airborne phase of flight." From the standpoint of senior decision makers, such high-level conclusions represent the ultimate payoff from vast amounts of data collection and analysis. To researchers and scholars, however, they beg more detailed questions. Are taxi-out times in the US high everywhere or are the results skewed by a few highly congested airports? Similarly, is airborne inefficiency fairly constant across space and time, or are there pronounced patterns of variation and, if so, what are they? Most importantly, what are the more important causes of flight inefficiency, and how can their contributions be quantified? Answers to such questions, in addition to satisfying basic curiosity, may be equally or more relevant to the practical project of improving system performance than the macro comparisons. For example, it may be far easier to import best practices from one region of a single air navigation system to another than to do so across systems.

2. LITERATURE REVIEW

The en route phase of a flight is defined as the portion between a 40 nautical-mile circular boundary around the departure airport (D40) and a 100 nautical-mile circular boundary around the arrival airport (A100). This definition is intended to exclude the portions of the flight path that are strongly influenced by terminal operations. Horizontal en route inefficiency, which evaluates actual flight trajectories against a benchmark trajectory, has received considerable attention in the open literature. Ref. [1] calculates the horizontal inefficiency based on the extra distance flown in the en route phase with respect to an ideal distance known as "achieved distance", which represents the average of how much further the flight has gotten from the origin and how much closer it has gotten to the destination over the en route portion of the flight *(2) - (3)*. This method, instead of choosing great circle distance between OD airport as the benchmark, excludes the effect of terminal inefficiency. Equation (1) explicitly expresses the definition, where HIE is the horizontal inefficiency of a flight, A is the actual flown distance, and H is the achieved distance.

$$HIE = \frac{A-H}{H} \tag{1}$$

The overall en route inefficiency can be further decomposed into two parts: direct route (DIR) extension inefficiency and terminal (TMA) extension inefficiency. While the first component is primarily driven by the efficiency of the path between the actual terminal area entry and exit points, the TMA extension inefficiency reflects the inefficiency that derives from the location of these points, which are usually not on the great circle route between the origin and destination. The decomposition can be written as:

$$HIE = DIR + TMA = \frac{A-D}{H} + \frac{D-H}{H} \tag{2}$$

Where DIR is the direct route extension inefficiency, TMA is the terminal extension inefficiency, and D is the great circle distance from the exit point to the entry point.

Based on this methodology, the US-Europe Performance Report *(1)* compares en route inefficiency in the US and Europe. Both US and Europe have been on a downward trend for en route inefficiency, but US in general in general is more efficient than Europe – in 2013, the flight inefficiency calculated from equation (2) is 2.7% for the US and 2.9% for Europe. The report documents certain patterns both for US and Europe. It suggests that flights to New York and Florida area are systematically more inefficient, mostly due to the avoidance of special use airspace (SUA) and long transcontinental operations. In Europe, the implementation of free route airspace (FRA) improves the en route efficiency significantly, especially for flights through those areas.

However, one major criticism of the metric is the selection of the "achieved distance" as the benchmark distance. Although it takes into account the deviation between exit/entry points with a direct route from origin to the destination, it is not an "optimum" trajectory distance for most flight operations, when considering meteorological conditions. Calvo et al. *(4)* propose the fuel efficiency as the metric to evaluate flight efficiency. Instead of using absolute distance, they calculate the inefficiency based on the additional fuel burn of the actual trajectory over the great circle trajectory between the exit and entry points. Although this metric performs quite differently from the route extension metric for some flights, these two are highly correlated most of the time. Therefore, route extension metric, while not perfect, has the virtue of simplicity and appears to correlate well with more refined metrics.

3. METHODOLOGY

Our aim is to investigate variation in the en route inefficiency metric for US flights. We focus on factors related to origin and destination airports, season, and flight distance. We apply linear regression techniques to explore and quantify factors that potentially impact flight horizontal en-route inefficiency. We are more interested in broad patterns than the specific circumstances that affect the inefficiency of individual flights.

3.1 Descriptive Data Analysis

We obtained the flight level performance data from the Enhanced Traffic Management System (ETMS) of FAA. The data cover around twelve million flights arriving at U.S. 34 core airports *(Appendix A)* from January 1st, 2013 to December 31st, 2014, in which 87% of total records are domestic flights and less 1% are diverted flights or missing records. Each record includes the origin/destination airports and departure/arrival time of a flight. Distance information driven by radar tracking data, such as actual flown distance, flight plan distance, great circle distance and achieved distance, is provided as well.

We limit our scope to flights that into and from the US 34 core airports, which represents most of US IFR flights *(1)*. After removing all international or diverted flights, we obtained our final dataset with six million records, which encompasses about 50% of total traffic in the ETMS database. Based on the data, we first apply Equation (1) to calculate the en route inefficiency for each flight, then we compare inefficiencies from the perspective of flight length, airport pair and season.

Figure 1 shows the average horizontal en route inefficiency for flights within each flight length category. There is no significant difference in inefficiency across all flight length groups between year 2013 and 2014. The average inefficiency across all flights is 3.413% for 2013, which is only 0.006% higher than 2014. Figure 1 shows that long-haul flights tend to be more efficient than short-haul flights, and that this is primarily the result of decreasing TMA extension inefficiency with distance. This suggests that excess distance from inefficient placement of entry and exit points is independent of great circle distance. On the other hand, DIR inefficiency is roughly independent of great circle distance, implying that the excess distance between entry and exit points compared to the great circle distance is roughly proportional to the great circle distance.

We calculate the monthly average horizontal inefficiencies for four representative routes, ATL – JFK, ATL – EWR, MSP – LAX and MSP – MIA, in 2013 to further investigate how terminals and seasons affect flight en route inefficiency. Boxplots of inefficiencies across months for those pairs are shown in Figure 2 and Figure 3. All of the plots reveal that inefficiency is skewed to the right. The skew is most pronounced for city pairs that are relatively efficient, because the left tail is bounded by zero.

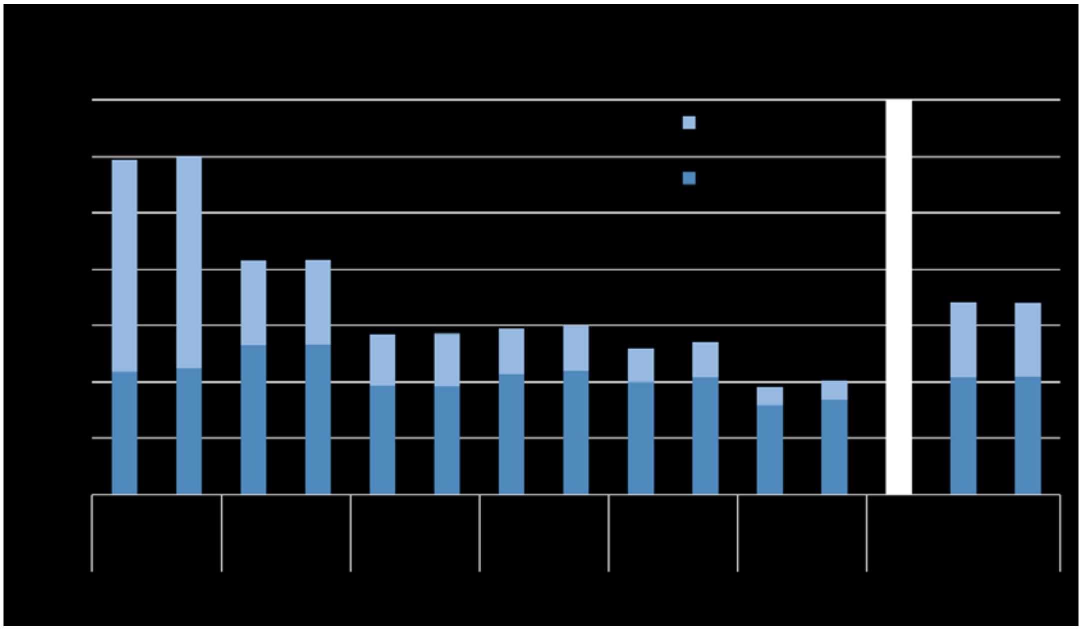


Figure 1 Summary of horizontal en route inefficiency

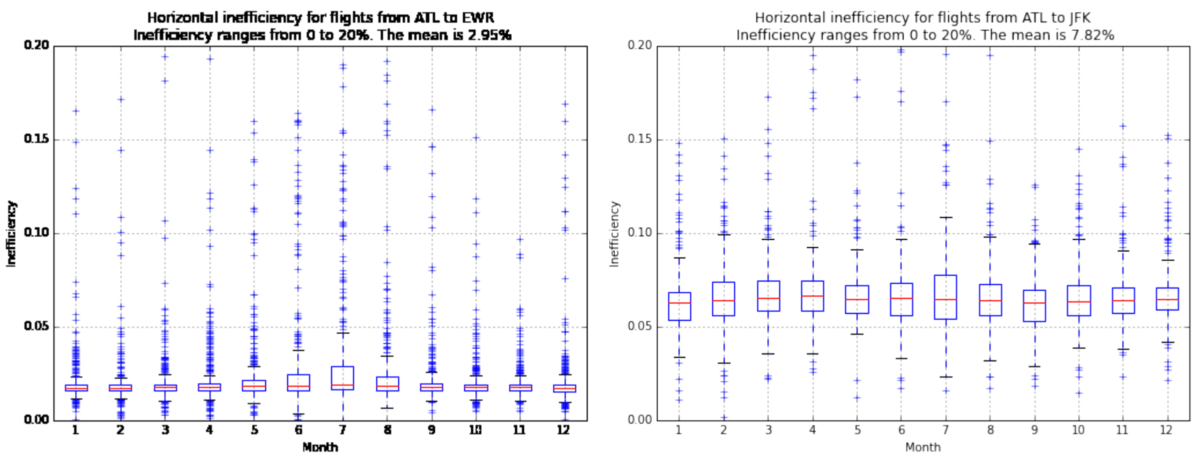


Figure 2 Horizontal inefficiency for representative airport pairs

Figure 2 illustrates the case for flights from ATL to JFK and from ATL to EWR. The average inefficiency across the whole year is 7.82% for flights to JFK, and is 2.95% for flights to EWR. Meanwhile, flights to JFK demonstrates higher variations than EWR. Since both pairs have the same origin airport and similar route structure, terminal congestion is likely to be a significant factor contributing to en route inefficiency. A possible explanation is that flights to JFK must circumvent traffic into the rest of the New York metroplex, while EWR, since it is on the southern edge of the region, is more accessible from the south. The second group of routes, shown in Figure 3, which have similar flight length, reveals the impact of season. Both figures indicate that flights from May to August are more inefficient than the other months, especially for flights from MSP to MIA. This suggests that convective weather, which is more frequent during summer season and along the south coast, increases the overall flight en route inefficiency.

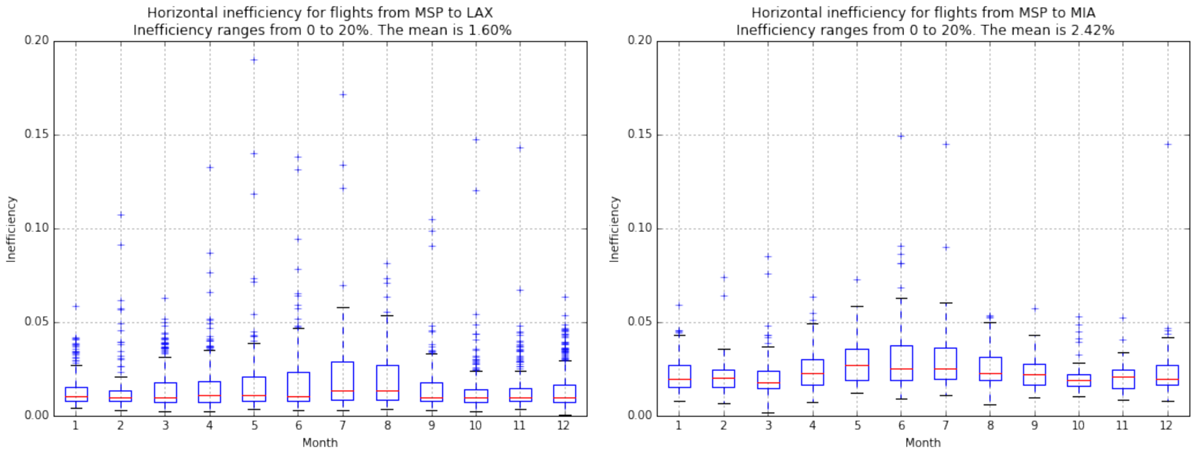


Figure 3 Horizontal inefficiency for representative airport pairs

3.2 Causal Analysis of Flight Inefficiency

To quantify the contributions of different causal factors to flight inefficiency, we employ two different approached, both of which are based on the concept of nominal routes. Nominal routes for a given airport origin and destination are identified first by clustering trajectories using dimension reduction and the DBSCAN algorithm, and then solving the 1-median problem to find a representative trajectory for each cluster. We term these representative trajectories as nominal routes. An example for one origin-destination pair, Houston Intercontinental to Boston Logan, appears in Figure 4. It considerable variation in the inefficiency for the different clusters, from less than 2% to nearly 9%, and that three clusters, colored red, green, and purple, account for the vast majority of flights.

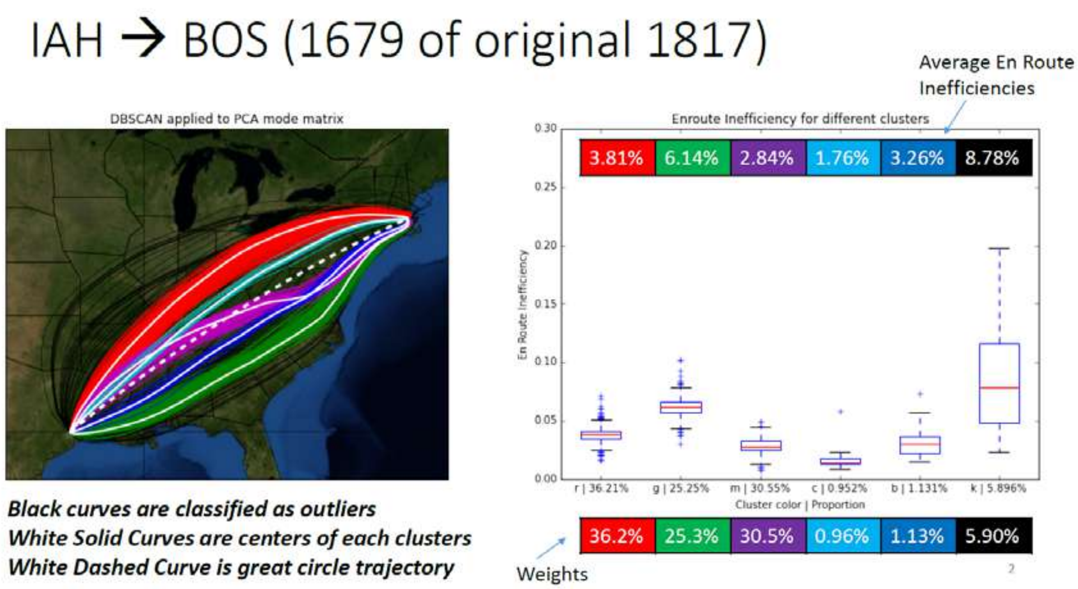


Figure 4 Example of Trajectory Clusters and Nominal Routes

Using the nominal routes, we employed two different methodologies for assessing contribution of different causal factors to en route inefficiency. For purposes of illustration, we will present these methodologies in the context of two causal factors—convective weather and rain. The first method, termed the composite approach, is shown in Figure 5. In this approach, the nominal routes are used as a basis for characterizing weather exposure for a given flight based on its departure time, which, along with other relevant factors such as whether a flight takes place during a busy hour of the day, is used as an independent variable in a regression model for the en route inefficiency of the flight. (Note that the composite measure is used instead of a flight-specific factor because the flight may have taken a longer trajectory to avoid the weather.)

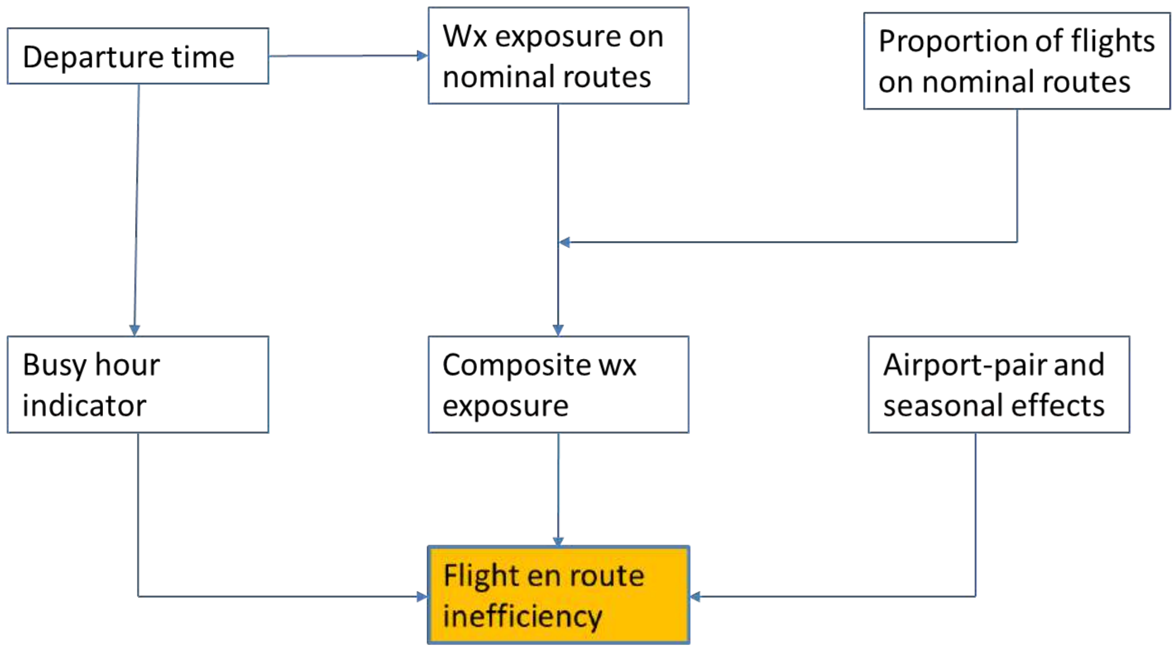


Figure 5 Composite Method

In the other approach, shown in Figure 6, route selection is explicitly modeled, using multinominal logit. The inefficiency impact is weather is depicted as a two-stage process, where the first stage is related to the cluster that is chosen for a given flight and the second stage models flight inefficiency in a manner that takes the cluster into account.

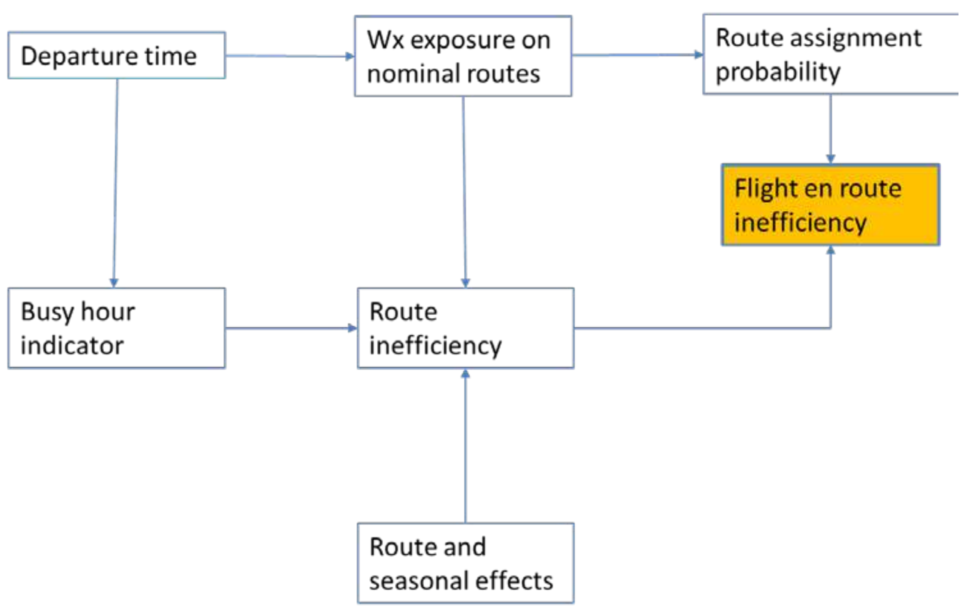


Figure 6 Route-Specific Method

For either approach, the final step is to use the estimated models to predict what inefficiency would be if there were no convective weather or rain present. The difference in inefficiency between the scenarios with the observed weather and the no weather is the contribution of weather to en route inefficiency.

4. RESULTS

Results of the analysis for several representative origin-destination pairs appear in Table 1. The composite method yields higher estimated of the weather contribution to en route inefficiency, which range from 14 to 27% of total inefficiency, with an average of about 20%. The route-specific method yields contributions between 6 and 16%, with an average of 9%. Efforts to understand these differences and if possible reconcile the divergent estimates are ongoing.

Composite Method

Airport Pair	Average Inefficiency with Convective Weather	Average Inefficiency without Convective Weather	**Percentage Change**
BOS_IAH	2.58%	2.03%	**26.94%**
DCA_ORD	4.20%	3.66%	**14.50%**
FLL_JFK	3.59%	2.71%	**32.48%**
IAH_BOS	4.37%	3.75%	**16.34%**
JFK_FLL	3.05%	2.15%	**41.72%**
JFK_LAX	2.05%	1.70%	**20.19%**
LAX_JFK	2.43%	2.07%	**17.30%**
ORD_DCA	4.11%	3.57%	**15.10%**
Overall	*3.10%*	*2.59%*	***19.95%***

Route-specific Method

Airport Pair	Average Inefficiency with Convective Weather	Average Inefficiency without Convective Weather	**Percentage Change**
BOS_IAH	2.58%	2.23%	**15.72%**
DCA_ORD	4.20%	3.82%	**9.70%**
FLL_JFK	3.59%	3.21%	**11.75%**
IAH_BOS	4.37%	4.14%	**5.51%**
JFK_FLL	3.05%	2.75%	**10.79%**
JFK_LAX	2.05%	1.91%	**6.96%**
LAX_JFK	2.43%	2.28%	**6.46%**
ORD_DCA	4.11%	3.74%	**9.67%**
Overall	*3.10%*	*2.85%*	***8.95%***

59

5. CONCLUSIONS

In this research, we present a method for quantifying to contribution of different causal factors to en route flight inefficiency, use weather factors as an example. Two different approaches, both based on characterizing the weather factors on nominal routes identifies using clustering of individual trajectories, are presented and compared. The results of the two approaches are somewhat different, but generally suggest that weather accounts for about 10-20% of en route inefficiency in the origin-destination pairs investigated to date. Put another way, flights fly in these pairs fly an excess ground distance of about 3.1% compared to a perfectly efficient great circle route; without adverse weather this would become 2.6 or 2.9% depending on the method used to estimate it.

In addition to refining these methods in order to get closer agreement between them, future research should focus on other causal factors. For example, traffic management actions such as increase mandatory spacing between aircraft to meter demand, are expected to increase en route inefficiency. Also, the effects of winds are important: flights may use a trajectory with a longer ground distance because it has more favorable winds. Ultimately, this research will lead to an understanding of the causal factors leading en route inefficiency comparable to our understanding of the causes of flight delay, which is currently significantly greater.

REFERENCES

1. Performance Review Commission and FAA-ATO, 2013 Comparison of Air Traffic Management-Related Operational Performance: U.S./Europe. *European Commission, U.S. Department of Transportation*, 2014.
2. EUROCONTROL, Horizontal flight efficiency, achieved distances. *Technical report, EUROCONTROL*, 2013. https://www.eurocontrol.int/sites/default/files/content/documents/single-sky/pru/news-related/2013-05-08-slides-workshop-achieved-distance.pdf. Accessed November 9, 2016.
3. Fuller I, Hustache JC, Kettunen T. Enhanced Flight Efficiency Indicators. *Eurocontrol Experimental Centre*, EEC/SEE/2004/011. 2004.
4. Calvo, E., Cordero, J. M., D'Alto, L., López-Leonés, J., Vilaplana, M., & La Civita, M. A New Method to Validate the Route Extension Metric against Fuel Efficiency. In *Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)*, Lisbon, Portugal, 2015
5. Sridhar, B., Chen, N.Y., Hok, K.N., Rodionova, O., Delahaye, D. and Linke, F. Strategic Planning of Efficient Oceanic Flights. In *Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)*, Lisbon, Portugal, 2015
6. McNally, D., Sheth, K., Gong, C., Sahlman, S., Hinton, S., Lee, C., Sterenchuk, M. and Shih, F.T. Dynamic Weather Routes: Two Years of Operational Testing at American Airlines. In *Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)*, Lisbon, Portugal, 2015
7. Meyers, T.A., Klopfenstein, M., Mintzer, J., Wilmouth, G. and Sud, V. A Preliminary Analysis of the Impact of Miles-in-Trail Restrictions on NAS Flight Operations. In *Sixth USA/Europe Air Traffic Management Research and Development Seminar (ATM2005)*, Baltimore, MD, 2005
8. Ostwald, P., Topiwala, T. and DeArmon, J. The Miles-in-Trail Impact Assessment Capability. In *AIAA Aviation Technology, Integration and Operations Conference (ATIO)*, Wichita, KS, 2006.
9. Sheth, K., Gutierrez-Nolasco, S., and Petersen, J., Analysis and Modeling of Miles-in-Trail Restrictions in the National Airspace System, In *AIAA Aviation Technology, Integration and Operations Conference (ATIO)*, Los Angeles, CA, 2013.

Multimodal Transportation Services

TB1: IP Methods in Air Traffic Control

Thursday 1:00 – 2:30 PM

Session Chair: Michael Ball

1:00 Optimizing the Slot Allocation on a Network of Airports
[1]Paola Pellegrini, [2]Tatjana Bolić, [2]Lorenzo Castelli, [3]Raffaele Pesenti*
[1]IFSTTAR, [2]Università degli Studi di Trieste, [3]Università Ca' Foscari di Venezia

1:30 Greedy Policies for a Dynamic Stochastic Transportation Problem, and an Application to Air Traffic Management
Alexander Estes, Michael Ball*
University of Maryland

2:00 A Mechanism for Auctioning Airport Landing Slots with Explicit Valuation of Congestion
[1]Michael Ball, [1]Alex Estes, [2]Mark Hansen, [2]Yulin Liu*
[1]University of Maryland, [2]University of California-Berkeley

Optimizing the slot allocation on a network of airports

Paola Pellegrini[1], Tatjana Bolić[2], Lorenzo Castelli[2], Raffaele Pesenti[3]

[1] Univ. Lille Nord de France – IFSTTAR, COSYS, LEOST, Villeneuve d'Ascq, France
paola.pellegrini@ifsttar.fr
[2] Dipartimento di Ingegneria e Architettura, Università degli Studi di Trieste, Trieste, Italy
tbolic@units.it and castelli@units.it
[3] Dipartimento di Mangement, Università Ca' Foscari di Venezia, Venezia, Italy
pesenti@unive.it

1 Introduction

Demand for air transportation continues to increase, and airports often become bottlenecks due to their lack of capacity. Nowadays, most of the congested airports are classified as either *Level* 2 or *Level* 3 (IATA, 2015), at which the use of airport facilities must be coordinated by a facilitator or a coordinator, respectively. At *Level* 3 airports, the coordinators may impose *schedule displacements* (Pyrgiotis and Odoni, 2014) on the users' requests and manage the capacity through the so called *slots*, i.e., "permissions given for a planned operation to use the full range of airport infrastructure necessary to arrive or depart at a Level 3 airport on a specific date and time" (IATA, 2015). In Europe, the slot allocation process consists of two main steps: primary slot allocation, and slot exchanges and transfers. In the first step, the slot allocation is performed on an airport-by-airport basis. The requests with historical precedence (*grandfather rights*) are allocated first. Then, 50% of the remaining slots are assigned to *new entrant* users. The rest is distributed in a non-discriminatory manner to all other users. Therefore the slots users receive are the outcome of several local allocations, and as such they may include slots that make the fleet rotation impracticable, or result in an undesirable schedule. The next step takes place at the IATA conference, where users aim to adjust the obtained schedule displacement through negotiations and slot exchanges. After the IATA conference, negotiations may last until shortly before the operations. At *Level* 2 airports, the facilitator may only propose schedule displacements to airspace users, using permissions similar to slots. In the rest of the paper, for the ease of terminology, we will refer to both coordination and facilitation as slot allocation.

This work shows the benefits of optimally coordinating the capacity management of airports in Europe by simultaneously allocating slots in all the Level 3 and Level 2 airports. This is achieved through SOSTA, an integer linear programming (ILP) formulation for the Simultaneous Optimization of the SloT Allocation for a network of airports. SOSTA reproduces the current regulations and best practices, and minimizes the number of missed allocations and the total schedule displacement cost, in lexicographic order. It could be used to partially replace and shorten the current (lengthy) slot allocation process. In fact, a simultaneous slot allocation would significantly reduce the users' need to negotiate for schedule building and re-building, due to accepted and rejected slot requests. Moreover, as shown in the experiments, SOSTA may be a tool for scenario analysis through the impact assessment of possible regulation changes.

The weaknesses of the current slot allocation practice have been identified and discussed by various studies (see, e.g., the review by Zografos et al., 2016). In particular, the existing slot allocation process is highly inefficient because the management of its complexity (the allocation needs to comply with numerous criteria and rules) is still largely empirical. To mitigate these inefficiencies, Zografos et al. (2012) formulate an ILP model that implements EU regulations (and IATA guidelines) and solves the slot allocation for a single airport. SOSTA extends this model from one airport to a network of airports, by combining the findings in Castelli et al. (2012), Pellegrini et al. (2012a), and Corolli et al. (2014). These papers also address the simultaneous allocation of slots at different interconnected airports, but under a simplified setting.

2 SOSTA's key features

SOSTA is formulated to satisfy the EU Regulations and IATA Worldwide Slot Guidelines (IATA, 2015). In the following, the description of its main features. For a detailed analysis we refer the interested reader to Pellegrini et al. (2017).

Characteristics of slots. A slot is the right to use the airport facilities for take-off or landing within a time *interval*. An interval is characterised by a *start time* and a *length*. Lengths may vary across airports, but all intervals at the same airport have the same length. Slots must be allocated within a *maximum displacement*. We refer to slots associated to departure (arrival) movements as *departure (arrival) slots*. The departure and arrival slots associated with the same flight are named *coupled slots*. Finally, the two slots associated with two movements to be operated sequentially by the same aircraft performing two different flights are *turnaround slots*.

Decision variables. Two sets of binary variables are introduced to decide: i) whether a requested slot s has to be allocated ot not (*missed allocation*), and in the former case, ii) whether slot s is allocated to interval r or not.

Objective functions. SOSTA adopts a lexicographic approach: first, the number of missed allocations is minimized and then, within the remaining requests, the total cost of displacement is minimized.

Constraints. Besides imposing that each requested slot is allocated to exactly one interval unless it is subject to a missed allocation, the following constraints hold:

Table 1: Comparison of SOSTA and *Current*.

Presence of grandfather rights				Absence of grandfather rights			
Missed alloc.		Disp. cost		Missed alloc.		Disp. cost	
SOSTA	*Current*	SOSTA	*Current*	SOSTA	*Current*	SOSTA	*Current*
860	2586	10648	10634	334	1666	26981	37895

- *Grandfather rights*. Grandfather rights are upheld in the slot allocation process.
- *Capacity constraints*. Depending on the airport, one or more of the following capacity constraints is imposed: i) *Hourly capacity constraints.* The bound on the number of slots is given for periods of one hour, and it may be applied either on a rolling basis or sequentially. These capacity constraints may refer to arrival, departure or total slots. ii) *Interval capacity constraints.* The bound on the number of slots is given for each interval.
- *Couple constraints.* Either both coupled slot requests are accepted, or they are both subject to a missed allocation. If they are accepted, the demanded flight duration may be modified only within a predefined range.
- *Turnaround constraints.* The displacement of two turnaround slots can never imply a turnaround time shorter than a predefined value.

3 Experimental analysis

We present the results using SOSTA to allocate the slots requested for June 28^{th} 2013, the busiest day of the year, including all the European Level 3 (107) and Level 2 (79) airports. The model comprises about $145,000$ binary variables and $243,000$ constraints, and exploits real data: i) Airport capacities, from airport coordinators' websites and Demand Data Repository 2. The latter is a database containing airspace network and traffic data, developed and maintained by EUROCONTROL (the European Organisation for the Safety of Air Navigation). ii) Airspace users' slot requests, from Slot Coordinator, another EUROCONTROL's database.

First, we validate the model by comparing SOSTA's with the currently implemented allocation: the difference is only 6 slots out of the $32,665$ slots requested in the test day. SOSTA reaches optimality in 125 CPU seconds on a computer running Linux Ubuntu distribution version 14:04, using CPLEX 12:6 as integer linear programming solver, and exploiting eight Intel Xeon 3.5 Ghz processors with 128 GB RAM.

Second, we test SOSTA's behavior when a significant imbalance between demand and capacity exists. Specifically, we compare SOSTA's and the optimal allocation under the *Current* rules assuming a uniform 20% reduction of airport capacity across Europe. The left-hand side of Table 1 shows that the simultaneous slot allocation at all airports significantly outperforms the current allocation process: the number of missed allocations decreases by 67% and the total displacement cost remains almost unchanged. Since the minimization of the number of missed allocations is the first objective of SOSTA, it always chooses a solution with fewer missing allocations, even

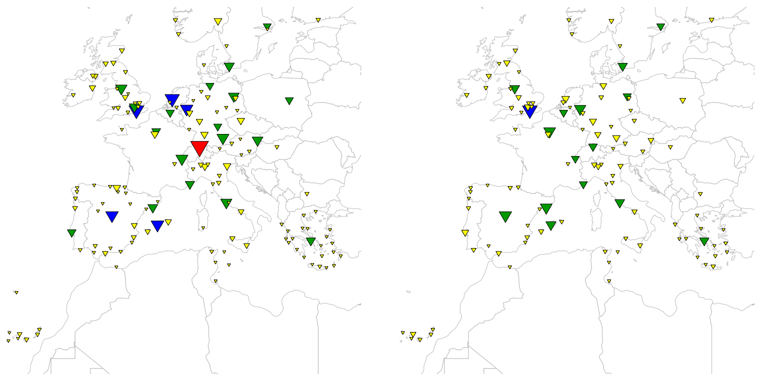


Figure 1: Airport-by-airport decrease of missed allocations passing from *Current* to SOSTA, when grandfather rights are either granted (left) or not (right).

if this may imply a higher displacement cost. SOSTA also shows good computational performances, as it just needs 140 CPU seconds to reach the optimality.

The right-hand side of Table 1 reports the results when the above comparison is repeated in absence of grandfather rights. Here SOSTA is used as a tool for scenario analysis. As expected, the results show that fewer missed allocations are necessary when no grandfather rights are granted. Moreover, SOSTA allows to further improve with respect to *Current* as it reduces the number of missed allocations by 80% and the total displacement cost by 29%. The computational performance worsens in the absence of grandfather rights, due to the elimination of several constraints and the consequent increase of the flexibility of the allocation: the optimality is reached after 19,107 CPU seconds. This corresponds to slightly more than one hour of wall-clock time on the computer used.

Figure 1 compares the number of *Current* and SOSTA missed allocations at each European airport, when grandfather rights are either granted (Figure 1-left) or not (Figure 1-right). SOSTA always allows a significant reduction of missed allocations: the larger the triangles' size, the larger the difference between *Current* and SOSTA . In particular, the triangles are red if this difference is greater than 75 missed allocations, blue if between 50 and 75, green if between 25 and 50, and yellow otherwise. While the difference is clearly dependent on the presence of grandfather rights, it appears that the largest differences are found at some busy (but not the busiest) airports: more than 50 missed allocations can be avoided by using SOSTA at Zürich, London Gatwick, Amsterdam Schiphol, Palma de Mallorca (the test day is a summer day), Düsseldorf and Madrid Barajas airports, when granting grandfather rights, and only at London Gatwick and Madrid Barajas airports, otherwise. Instead, in the busiest European airports this difference is more mitigated (always lower than 50 missed allocations): in London Heathrow, Paris Charles de Gaulle and Frankfurt the differences are, respectively, of 39, 31 and 18 missed allocations when granting grandfather rights, and 22, 47 and 20 otherwise.

4 Conclusions

In this work we described SOSTA, an ILP model that allows optimizing the slot allocation process in Europe respecting the current regulations and best practices. In the experimental analysis, we optimized the allocation of the busiest day of 2013. After validating SOSTA against the current practice, we assessed its performance in case of large imbalance between demand and capacity. Finally, we used it as a scenario analysis tool to observe the impact of the presence of grandfather rights. In all the tests performed, SOSTA outperformed our emulation of the current practice. In future research, we will consider further features of the slot allocation process, as the allocation of series of slots.

References

L. Castelli, P. Pellegrini, and R. Pesenti. Airport slot allocation in Europe: economic efficiency and fairness. *International Journal of Revenue Management*, 6(1/2):28–44, 2012.

L. Corolli, G. Lulli, and L. Ntaimo. The Time Slot Allocation Problem under Uncertain Capacity. *Transportation Research Part C - Emerging Technologies*, 46:16–29, 2014.

EUROCONTROL. PRR 2014. Performance Review Report - An Assessment of Air Traffic Management in Europe during the Calendar Year 2014. Brussels, Belgium, 2015.

IATA. Worldwide Slot Guidelines - 7. Technical Report 7, International Air Transport Association - Joint Slot Advisory Group, 2015.

P. Pellegrini, T. Bolić, L. Castelli, and R. Pesenti. SOSTA: an effective model for the Simultaneous Optimisation of airport SloT Allocation. *Transportation Research Part E - Logistics and Transportation Review*, 2017. To appear.

P. Pellegrini, L. Castelli, and R. Pesenti. Metaheuristic algorithms for the simultaneous slot allocation problem. *IET Intelligent Transport Systems*, 6(4):453–462, 2012a.

N. Pyrgiotis and A.R. Odoni. On the impact of scheduling limits: A case study at Newark Liberty International Airport. *Transportation Science*, 50(1):150–165, 2014.

K.G. Zografos, Y. Salouras, and M.A. Madas. Dealing with the Efficient Allocation of Scarce Resources at Congested Airports. *Transportation Research Part C - Emerging Technologies*, 21(1):244–256, 2012.

K.G. Zografos, M.A. Madas, and K.N.Androutsopoulos. Increasing airport capacity utilisation through optimum slot scheduling: review of current developments and identification of future needs. *Journal of Scheduling*, doi:10.1007/s10951-016-0496-7, 2016.

Greedy Policies for a Dynamic Stochastic Transportation Problem, and an Application to Air Traffic Management

Alexander Estes, Michael Ball
University of Maryland-College Park
aestes@math.umd.edu, mball@rhsmith.umd.edu

1 Introduction

The transportation problem is one of the most fundamental and well-known problems in combinatorial optimization. In this problem, there is a set of sources that supply some resource and a set of destinations that demand that resource. The goal is to match the supplied resources with the demand while minimizing cost. If there is an ordering on the pairs of sources and destinations, then a greedy algorithm for the transportation problem can be implemented by simply considering the pairs in order and assigning as much remaining resource as possible between the pair. Necessary and sufficient conditions for the optimality of this greedy algorithm are well-known (Hoffman 1963, Dietrich 1990, Adler et al. 1993). Specifically, the greedy algorithm will produce an optimal solution for all possible choices of supply and demand if and only the ordering of pairs is a *Monge sequence*.

We provide a formulation for a variant of the transportation problem which is dynamic in the sense that the problem occurs over several time periods, and is stochastic in the sense that new supply and demand arrives according to a random process. Spivey and Powell (2004) formulated a less general version of this problem and produced heuristics for its solution. We propose a class of greedy policies and provide sufficient and necessary conditions for the optimality of these policies.

This work can be applied to the problem of planning a type of air traffic management initiative called a ground delay program. In a ground delay program, some flights scheduled to travel to some airport are delayed on the ground in order to prevent excess congestion from forming in the air near the destination airport. Richetta and Odoni (1993) provided the first formulation of the problem of planning a ground delay program at an airport, which they called the single airport ground holding problem (SAGHP). Ball et al. (2010) proposed a variant of this problem and showed that a greedy algorithm provided an optimal solution. It was later shown that this variant may be reformulated as a transportation problem, and the optimality of the greedy algorithm can be explained by known properties of greedy algorithms for the transportation problem (Glover and Ball 2013). We propose a new dynamic formulation for planning ground delay programs, and we provide an optimal dynamic policy for this problem.

2 A Policy for Dynamic Ground Delay Program Control

In our formulation for the problem of planning a ground delay program, flights are dynamically assigned arrival times. This problem takes place over T discrete time periods. At the beginning, there are a set of flights $\mathcal{A}$ that wish to travel to an airport and a set of available slots $\mathcal{B}$ at this airport. A slot represents a set of resources required to accommodate the arrival of a flight in a specific time period. Each slot b then has a corresponding time period τ_b^{slot}. A flight arriving at the airport would only be able to land if a slot were available at that time.

The problem is to decide in each time period which flights will be allowed to depart and which flights will not. We assume that each flight a has an original scheduled arrival time τ_a^{OETA} and it cannot depart earlier than this time period. Each flight a also has a fixed, known flight time τ_a^{fly}. In this problem, we do not allow a flight to depart unless there is an available slot that corresponds with the arrival of the flight. That is, in order for flight a to depart in time period t then there must be an available slot b such that

$$t + \tau_a^{\text{fly}} = \tau_b^{\text{slot}}$$

If there is such a slot and the flight is allowed to depart, then the slot is marked unavailable. If a flight is not allowed to depart at its original scheduled arrival time then that flight must be delayed. The amount of delay experienced by that flight will continue to lengthen until the flight is allowed to depart. The goal of the problem is to minimize the total delay experienced by the flights. In each time period, some new flights or new slots may appear according to some random process. In order for a policy to be a successful policy for this problem, it must take into account the potential for new flights or slots to arrive.

We define a policy for this problem that we will refer to as the D-RBD policy. The D-RBD policy bears resemblance to the RBD algorithm described in Ball et al. (2010). However, D-RBD is a dynamic policy that produces assignments in each time period, while RBD is an algorithm for preallocating flights to slots at the beginning of the planning period for the ground delay program. In each time period t, the D-RBD policy examines all flights whose original ETA is at least t. This is exactly the set of flights that could depart if there were available slots. These flights are sorted according to their flight times, in decreasing order. The D-RBD policy then considers each flight in order, and each flight is allowed to depart if there is an available slot for that flight.

Theorem 1. *The D-RBD policy always achieves the minimum total delay of any policy.*

As far as we are aware, this is the first dynamic policy proposed for ground delay program planning that achieves the optimal solution without requiring the distribution of the random process that determines when new slots and new flights will appear. Our formulation of the ground delay program planning problem can be viewed as a special case of a more general type of dynamic, stochastic, transportation problem, and Theorem 1 can be derived from a result that we can prove in the more general context.

3 Formulation of a Dynamic Stochastic Transportation Problem

We formulate a variant of the transportation problem that includes dynamic and stochastic elements. There are T discrete time periods. As in the standard transportation problem, there is a set of sources $\mathcal{A}$ and a set of destinations $\mathcal{B}$. In each time period, new supply may appear at some sources and new demand may appear at some destinations. This is determined by some random process, which we will refer to as Ω. We will assume that whenever new supply appears, it appears at a new source that has not yet received supply, and we can assume likewise that new demand always appears at a new destination. In fact, we can make this assumption without any loss of generality. We will let $\mathcal{A}_t$ and $\mathcal{B}_t$ be the set of new sources receiving supplies in time period t, and we similarly define $\mathcal{B}_t$ to be the set of new destinations receiving demands in time period t. The state of our system in a time period is described by the supplies and demands that remain. We let $s_{t,a}^{\text{rem}}(\omega;\pi)$ be the amount of supply remaining at source a at the beginning of time period t in instance ω of the random process under policy π. We similarly define $d_{t,b}^{\text{rem}}(\omega;\pi)$ for demand remaining at destination b.

In each time period, the decision is how much supply to send from each source to meet demand at each destination. Let $x_{t,a,b}(\omega;\pi)$ be the amount of supply sent by the policy π from source a to meet demand in destination b in time period t of the instance ω. The amount sent from a to b is not allowed to exceed the supply remaining at a nor should it exceed the demand remaining at b. Thus, we have a constraint

$$x_{t,a,b}(\omega;\pi) \leq \min\{s_{t,a}^{\text{rem}}(\omega;\pi), d_{t,b}^{\text{rem}}(\omega;\pi)\}$$

We will refer to a pair of a source and a destination as an arc. Some arcs may be feasible while others may not, and the feasibility of an arc may change over time. We will let $\mathcal{F}_t$ be the set of feasible arcs at time t. Any movement of supply to meet demand incurs a cost. We let $c_{t,a,b}$ to be the cost of allocating one unit of supply from source a to satisfy one unit of demand from destination b in time t. These costs are assumed to be known. The total cost of the assignment made by policy π in time period t of instance ω is then given by

$$C_t(\omega;\pi) := \sum_{(a,b)\in\mathcal{F}_t} c_{t,a,b} x_{t,a,b}(\omega;\pi)$$

After the cost is incurred, the supply and demand are both removed, which results in the post-decision state. Let $s^x_{t,a}(\omega;\pi)$ be the amount of supply at source a that remains after the allocations are made in time t in instance ω under policy π. Let $d^x_{t,b}(\omega;\pi)$ be defined similarly for the amount of demand at destination b. Then, our post-decision state is given by:

$$\begin{aligned} s^x_{t,a}(\omega;\pi) &= s^{\text{rem}}_{t,a}(\omega;\pi) - \sum_{b\in\mathcal{B}} x_{t,a,b}(\omega;\pi), \\ d^x_{t,b}(\omega;\pi) &= d^{\text{rem}}_{t,b}(\omega;\pi) - \sum_{a\in\mathcal{A}} x_{t,a,b}(\omega;\pi). \end{aligned}$$

After these allocations are made, the problem proceeds to the next time period, and new supplies and demands are observed. Let $s^{\text{new}}_{t,a}(\omega)$ and $d^{\text{new}}_{t,b}(\omega)$ be the amount of supply and demand appearing at the beginning of time t in instance ω. The state update is then given by

$$\begin{aligned} s^{\text{rem}}_{t+1,a}(\omega;\pi) &= s^{\text{new}}_{t+1,a}(\omega) && \text{for } a \in \mathcal{A}_{t+1} \\ d^{\text{rem}}_{t+1,b}(\omega;\pi) &= d^{\text{new}}_{t+1,b}(\omega) && \text{for } b \in \mathcal{B}_{t+1} \\ s^{\text{rem}}_{t+1,a}(\omega;\pi) &= s^{\text{x}}_{t,a}(\omega;\pi) && \text{for } a \in \mathcal{A}_0 \cup ... \cup \mathcal{A}_t \\ d^{\text{rem}}_{t+1,b}(\omega;\pi) &= d^{\text{x}}_{t,b}(\omega;\pi) && \text{for } b \ \in \mathcal{B}_0 \cup ... \cup \mathcal{B}_t \end{aligned}$$

We will say that a policy produces a feasible allocation if the policy allocates every unit of supply to a unit of demand. The goal of the problem is to identify the policy that will produce the minimum cost feasible allocation.

3.1 Deterministic Problem

Let us consider a problem where it is known which instance ω of the random arrival process will occur. We provide a formulation for the resulting problem as follows. Define $\mathcal{F}^*$ to be the set of all feasible arcs. That is,

$$\mathcal{F}^* := \mathcal{F}_0 \cup ... \cup \mathcal{F}_T$$

Under the assumption that knowledge of when our resources will arrive is available, then any optimal policy would make all allocations at the minimum cost time. This allows us to omit the time that the allocations occur from our decision variables. We will then let $x_{a,b}$ be a decision variable that represents the amount of resource from source a that is used to satisfy demand at destination b. The corresponding cost would be given by

$$c^*_{a,b} = \min\{c_{t,a,b} : (a,b) \in \mathcal{F}_t, t \in \{0, ..., T\}\}.$$

The resulting problem is given by:

$$\min \sum_{(a,b)\in\mathcal{F}^*} c^*_{a,b} x_{a,b}$$

such that:

$$\begin{aligned} \sum_{b:(a,b)\in\mathcal{F}^*} x_{a,b} &= s^{\text{new}}_{t,a}(\omega) && \forall t \in \{0, ..., T\}, a \in \mathcal{A}_t, \\ \sum_{a:(a,b)\in\mathcal{F}^*} x_{a,b} &= d^{\text{new}}_{t,b}(\omega) && \forall t \in \{0, ..., T\}, b \in \mathcal{B}_t. \end{aligned}$$

Note that this problem is a standard transportation problem. We will refer to this problem as the deterministic problem corresponding to instance ω. If a policy achieves the value of the deterministic problem for any instance ω of any arrival process Ω, then we will say that the policy is *oracle-optimal*. Given some ordering $\prec$ of the feasible arcs $\mathcal{F}^*$, a greedy algorithm can be implemented by considering each arc in order and allocating as much remaining supply along that arc as possible. We will use the notation that $x^*_{a,b}(\omega; \prec)$ is the amount of resource allocated on the arc (a, b) in the greedy solution using ordering $\prec$ for the deterministic problem corresponding to instance ω.

4 Subset Greedy Policies

We define a class of policies for the dynamic stochastic transportation problem as follows. In time period t, our policy considers a subset U_t of the set of feasible arcs $\mathcal{F}_t$. The policy orders these arcs according to an ordering $\prec_t$. Then, the policy considers the arcs in order and allocates as much remaining resources as possible. Such a policy is characterized by the sequence of subsets $U_0, U_1, ..., U_T$ and the sequence of corresponding orderings $\prec_0, \prec_1, ..., \prec_T$. In order to simplify notation, if we have an ordering $\prec$ on some set S and we have subsets $V_1, V_2 \subseteq S$ then we will say that $V_1 \prec V_2$ if $v_1 \prec v_2$ for any $v_1 \in V_1$ and $v_2 \in V_2$. We show that for any subset-greedy policy, there is an equivalent policy with disjoint subsets.

Lemma 1. *Let there be a subset-greedy policy π with subsets $U_0, U_1, ..., U_T$ and corresponding orderings $\prec_0, \prec_1, ..., \prec_T$. Let*

$$U'_t = U_t \setminus \left(\bigcup_{t'=0}^{t-1} U_{t'} \right)$$

and let $\prec'_t$ be the restriction of $\prec_t$ to the set U'_t. Then the subset-greedy policy π' with subsets $U'_0, U'_1, ..., U'_T$ and ordering $\prec'_0, \prec'_1, ..., \prec'_T$ is equivalent to the policy π.

As long as every feasible arc appears in at least one of the subsets $U_0, ..., U_T$, we can observe that the assignment made by a subset greedy policy will produce the same assignments as a greedy solution to the deterministic problem, although the subset greedy policy will not necessarily make them at the optimal time.

Corollary 1. *Let there be a subset-greedy policy π with subsets $U_0, ..., U_T$ with corresponding orderings $\prec_1, ..., \prec_T$. If $U_0, ..., U_T$ form a partition of $\mathcal{F}^*$, then there exists an ordering $\prec^*$ on $\mathcal{F}^*$ such that*

$$\sum_{t=0}^{T} x_{t,a,b}(\omega; \pi) = x^*_{a,b}(\omega; \prec^*)$$

for any $a \in \mathcal{A}$, $b \in \mathcal{B}$, and for instance ω of any arrival process Ω. In particular, the ordering $\prec^*$ on $\mathcal{F}^*$ is defined such that:

1. *$U_t \prec^* U_{t'}$ for any times τ, τ' with $t < t'$,*
2. *for $u_1, u_2 \in U_t$ then $u_1 \prec^* u_2$ if and only if $u_1 \prec_t u_2$.*

If the policy does make all of its assignments at the minimum cost times, then it will achieve the same cost as the corresponding greedy solution to the deterministic problem. This allows us to use known conditions for optimality of greedy solutions to transportation problems to produce sufficient and necessary conditions for the oracle-optimality of a subset-greedy policy. Let $\mathcal{M}_t$ be the subset of feasible arcs $\mathcal{F}_t$ in time period t such that the minimum cost for moving resource along these arcs is reached in time period t.

Theorem 2. *Let π be a subset-greedy policy with disjoint subsets $U_0, U_1, ..., U_T$ and corresponding orderings $\prec_0, \prec_1, ..., \prec_T$. Let $\prec^*$ be the corresponding ordering on $\mathcal{F}^*$ as described in . Then π is oracle-optimal if and only if $U_t \subseteq \mathcal{M}_t$ for all t, the subsets $U_0, U_1, ..., U_T$ form a partition of $\mathcal{F}_t$, and $\prec^*$ produces a Monge sequence for the deterministic problem.*

A subset-greedy policy is oracle-optimal if and only if it always makes allocations at the minimum cost time, it considers all feasible allocations, and it corresponds to a Monge sequence of the deterministic problem.

Our result for the variant of the SAGHP is a special case of our result for the dynamic stochastic transportation problem. We can show that the former problem may be reduced to the latter problem, and that the D-RBD policy is a special case of a subset greedy policy. Then we can show that the requirements of Theorem 2 are satisfied, which implies the optimality results given in Theorem 1.

5 Conclusion and Further Work

We provide a formulation for a new transportation problem to a setting that is both dynamic and stochastic, and we describe a type of greedy policy for this problem. We provide necessary and sufficient conditions under which this greedy policy will achieve the same solution that would be achieved if we had complete knowledge of the stochastic process. Using these conditions, we can prove the optimality of a dynamic policy for ground delay program planning Ball et al. (2010).

Since we provided necessary and sufficient conditions for the oracle-optimality of a certain class of policies, our work naturally provides sufficient conditions for the existence of an oracle-optimal policy for the dynamic stochastic transportation problem. However, it seems plausible that some instances of the dynamic stochastic transportation problem may have oracle-optimal policies that are not subset greedy policies. Further work could provide a more complete characterization of the conditions under which oracle-optimal policies exist. We also do not provide a practical method for checking whether or not our conditions are satisfied. Future research could produce an algorithm that would determine whether or not an oracle-optimal subset greedy policy exists for a dynamic stochastic transportation problem.

While Ball et al. (2010) showed that the RBD algorithm was optimal in terms of minimizing delay, they also showed that this algorithm can produce allocations that are not very equitable. Our policy would likely suffer from the same weakness. It may be possible to extend the problem to include equity considerations while still preserving the properties of the original problem that allow for a simple optimal policy.

References

Adler I, Hoffman AJ, Shamir R (1993) Monge and feasibility sequences in general flow problems. *Discrete Applied Mathematics* 44(1):21–38.

Ball MO, Hoffman R, Mukherjee A (2010) Ground delay program planning under uncertainty based on the ration-by-distance principle. *Transportation Science* 44(1):1–14.

Dietrich BL (1990) Monge sequences, antimatroids, and the transportation problem with forbidden arcs. *Linear Algebra and Its Applications* 139:133–145.

Glover CN, Ball MO (2013) Sparse monge matrices arising from scheduling problems. *Operations Research Letters* 41(3):246–248.

Hoffman A (1963) On simple linear programming problems. *Proceedings of Symposia in Pure Mathematics*, volume 7, 317–327.

Richetta O, Odoni AR (1993) Solving optimally the static ground-holding policy problem in air traffic control. *Transportation Science* 27(3):228–238.

Spivey MZ, Powell WB (2004) The dynamic assignment problem. *Transportation Science* 38(4):399–419.

A mechanism for auctioning airport landing slots with explicit valuation of congestion

Michael O. Ball & Alex Estes
University of Maryland, College Park, MD 20742
Mark Hansen & Yulin Liu
University of California, Berkeley, CA 94720

January 5, 2017

The economic consequences of air transportation delays are well-known (e.g. Ball et al. 2010, Ferguson et al. 2013) and much research has been devoted to reducing them. One category of approach is to limit the demand placed on the underlying system. This is most often accomplished by defining a set of airport slots and requiring that any flight takeoff or landing operation use a designated slot. Such slots are nearly always allocated and reallocated to flight operators using various administrative rules. There is a growing body of research and specific societal efforts that investigate the merits and feasibility of using market-based approaches for such allocations. When one considers the fundamental nature of this problem, it becomes clear that the required market mechanism is some form of combinatorial auction (see Ball et al. 2006, Rassenti et al. 1982). A combinatorial auction solves the problem of finding prices to sell a diverse set of products where potential bidders/buyers place values on packages consisting of multiple products. In the airport setting a product is the right to conduct a landing or takeoff operation within a specific time window. Of course, a scheduled air carrier will typically desire multiple slots within a given time window and also slots from multiple time windows, hence the need for package bidding and a combinatorial auction.

A key input to defining an airport slot auction is the definition of time window widths and the number of slots available within each time window. The number of operations a runway can support depends on the sequence and mix of aircraft types, the efficiency and skill of the human operators (air traffic controllers and pilots), weather conditions as well as other factors. For these reasons, the system should be modeled as one in which expected congestion and delays increase as a function of the number of slots rather than one in which there is a hard limit on the number of operations. Viewed from this perspective one is then left with the challenge of defining the number of slots to assign to each time window.

There is a body of work that seeks to define appropriate slot levels (e.g. Churchill et al. 2012) Inevitably, any approach must evaluate the fundamental tradeoff: more slots lead to more scheduled operations at the expense of more congestion and delays. In a setting where a single airline offered all flights at an airport that airline should be able to properly evaluate this tradeoff. However, in a competitive environment, a single airline could decide to add one or two additional flights and only internalize the delay on those two flights. Yet, those additional flights would induce additional delay on several other flights (owned by other airlines). In fact, the failure of an airline to internalize all the delay costs it causes is a fundamental reason that slot controls are necessary. The premise of this paper is to define an auction mechanism where the valuation problem required of airline bidders is the same that would be required of an airline operating in non-competitive environment.

Specifically, under our model, a set of time windows, $t = 1, ..., T$ and a set of slot profiles $p \in \Gamma$ are defined, where associated with each $p \in \Gamma$ is a vector of slot limits: $(n_1(p), ..., n_T(p))$. We assume each flight operator/bidder is able to evaluate a value function $v_j(y_{j1}, ..., y_{jT}, p)$, which gives the value to that bidder of the slot vector $(y_{j1}, ..., y_{jT})$ under profile p, where y_{jt} equals the number of slots owned by flight operator j in time window t. The underlying assumption is that a slot profile p can be mapped by the flight operator to a level of congestion and delays and that a value can be placed on a slot vector within that environment.

We now investigate how a slot auction can be developed using these bidder valuation capabilities. A fundamental result in combinatorial auctions is that the use of Vickrey-Clark-Groves (VCG) prices (see Cramton et al. (2006) for background) are incentive compatible in the sense that a dominant strategy for bidders is to submit bids for packages that are equal to their internal valuation of those packages. It has been widely observed in the open literature that this result is quite robust relative to the nature of the winner determination problem solved.

In our framework there is a set flight operator/bidders, N, and each $j \in N$ can formulate bids, $x_j = (y_{j1}, ..., y_{jT}, p)$ and associate a value $v_j(x_j)$ with such bids. We denote by Φ_j the set of all possible bids of interest to j. Bidder j is responsible for defining a bid function $b_j(\mathbf{x}_j)$, which maps each $\mathbf{x}_j \in \Phi_j$ to a bid price. A key issue in auction design is the relationship between the bidder's private value function v_j and b_j. The VCG result referenced in the prior paragraph implies that that bidders are incentivized to set $b_j = v_j$. In general, the resources awarded by the auctioneer must satisfy a set of constraints, Ω. That is, the resources allocated to all bidders, $(\mathbf{x}_j)_{j \in N}$ must satisfy, $(\mathbf{x}_j)_{j \in N} \in \Omega$. The basic set of constraints in our setting would be that a single profile p be chosen and that the associated slot limits be respected, i.e. $\sum_j y_{jt} \leq n_t(p)$ for all t.

Allowing the bidding process to set the slot limits, brings to the forefront the potential for bidders to use the auction process to increase market power. Specifically, as an bidder's share of slots increases, that bidder gains market power allowing some degree of monopoly rents to be extracted. This would not be considered a desirable outcome. The auction design we contemplate not only allows market power to be gained by increasing a given bidder's slot holdings but also by reducing overall holdings. Because of this, we feel it is important to explicitly limit marker power within the auction design. A common measure of market concentration is the Herfindhal index (HHI): $\sum_{i \in N} \alpha_i^2$ where N is the set of competitors in the market and α_i is competitor i's fraction of total market share. To apply the HHI to our problem we would define:

$$\alpha_i \quad = \text{(flight operator } i \text{ slot holding)/(total number of slots)} \tag{1}$$

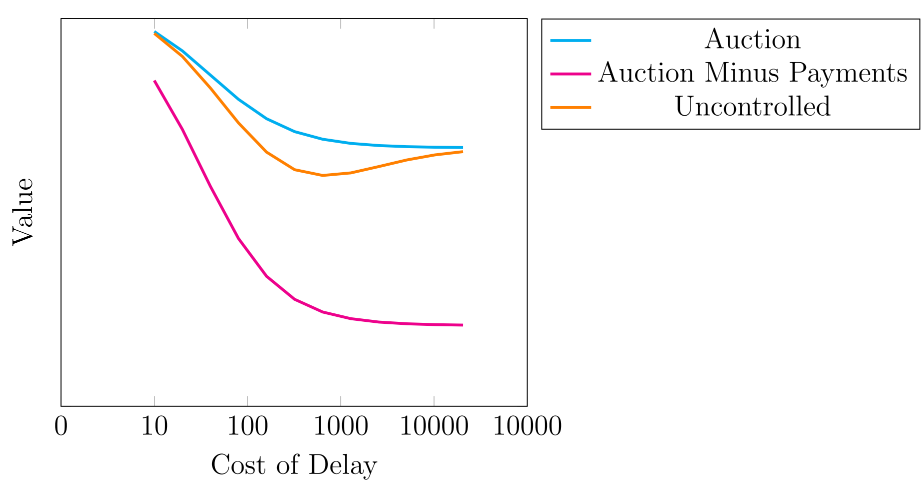
We propose to apply the HHI to slot holdings over two or more periods of the day. For example, if there were two busy periods during the day and say 2 to 4 time windows covered each of these busy periods then constraints could be developed that limited the HHI value for these two periods. Such constraints would be incorporated into Ω; appropriate linear constraints can be defined within a linear integer programming representation of the winner determination problem.

Thus, far we have developed a conceptual approach to a combinatorial airport slot auction that achieves the objectives we discussed at the outset. There are many further details to be developed in order to achieve a practical mechanism design. Our objective here is not to try to accomplish this. We contend that the concepts outlined, the basic VCG result as well as the existing knowledge base on combinatorial auctions, e.g. Cramton et al. (2006) imply that a such a mechanism could be developed. Rather, in this paper we seek to investigate other issues related to the viability of our concepts. Specifically, we consider the overall welfare gain from this approach as well as the distribution in consumer surplus among the parties involved. To accomplish this, we develop two models. The first is a stylized continuous approximation and the second is an integer programming model applied applied to a historical dataset.

1. Continuous Approximation

In our model we assume a set of identical flight operators. The flight operator value function increases with the number of slots it is given. The number of flight operations is proportional to the number of slots assigned, and these operations are distributed uniformly throughout the day. We compute flight delays using a deterministic queueing model. The value of a set of slots to a flight operator decreases linearly with the total delay. With this model we can find closed-form expressions for the total auction value as well as the VCG payments. Further, we can find a closed-form for expression for the total value of an "uncontrolled schedule", i.e. the flight schedule that would result in the absence of any slot controls. Figure 1 provides a graphical representation of these results.

Figure 1: Results from Continuous Approximation



The series 'Auction' shows the total value of the slots allocated in the auction, while the series 'Auction Minus Payments' displays this value minus the payments that the flight operators make in the auction. The series 'uncontrolled' shows the value that the flight operators derive in the uncontrolled setting. These computational results show the degree to which the total value of the schedule resulting from the auction exceeds the value of the schedule resulting in an uncontrolled setting. They also show (not surprisingly) that the flight operators will be worse off if they are required to participate in such an auction. Further results show the total delay reduction achieved (which is more valuable to passengers – and society – then to flight operators). These results indicate that such an auction is desirable from a public policy perspective but that gaining flight-operator buy-in will require explicit policy measures.

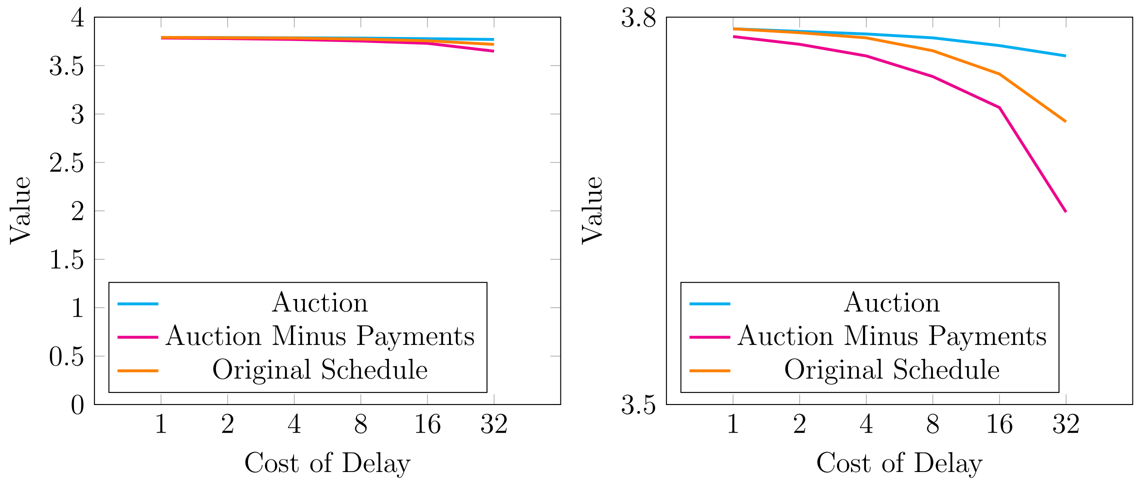
2. Integer Programming Model

Our integer programming model represents the winner determination problem for a specific set of bidder valuation functions. We start with a historical schedule and assume that this contains the desired schedule for each airline whose flights are included in the schedule. The model includes a vector of slot requests for each airline j as defined earlier, $(y_{j1}, ..., y_{jT})$ It also includes slot profile variables and constraints, insuring that a single profile is chosen and that the total number of slots assigned does not exceed the limits indicated in the profile. A set of flight variables and constraints are defined for each airline. Flights may be left as they are in the current schedule, moved to an alternate time period or removed from the schedule. Appropriate costs are associated with the second two options. An average flight delay is computed for each time window under each slot profile by applying an airport delay model (Mukherjee et al. 2005) to the possible profiles. The delay associated with a slot is applied to the flights arriving or departing in each slot and an appropriate delay cost is assigned to these flights. Thus, each airline's value function is defined through this set of constraints and costs. We solve this model to find the total value produced by the auction and can also compute the VCG payments. We can apply the same delay cost function to the original schedule to obtain the net value of that schedule (flight values minus delay costs). This procedure was carried out with several values of delay. The results are shown in Figure 2. The two plots in Figure 2 contain the same information. The series labeled 'Auction' shows the total value of the schedule produced by the auction. In order to generate the series labeled 'Auction Minus Payments,' we subtract the payments made by the flight operators from the value of the slots that they are allocated in the auction. The value of the original schedule is shown in the series labeled 'Original Schedule.' Similarly to the results of the continuous model, we see that the auction is able to produce a better allocation of slots than the original schedule. However, the payments that the flight operators are larger than the gain in value that the flight operators receive from the better allocation. Further analysis of these results provide specific information on delay reduction and change in schedule value for each airline.

3. Conclusions

This paper defines, at a conceptual level, a new combinatorial auction mechanism in which the congestion levels are set by the bidders within the auction mechanism. It evaluates this

Figure 2: Results from IP Model



concept for the challenge of auctioning airport landing slots. The results provide insight into the public policy challenges and approaches to overcoming these challenges are discussed. This approach is compared to recently developed administrative approaches, e.g. (Jacquillat and Odoni 2015).

References

Ball M, Barnhart C, Dresner M, Hansen M, Neels K, Odoni A, Peterson E, Sherry L, Trani AA, Zou B (2010) Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States. Technical report, NEXTOR Research Symposium, Washington DC.

Ball M, Donohue G, Hoffman K (2006) Auctions for the safe, efficient, and equitable allocation of airspace system resources. Cramton P, Shoham Y, Steinberg R, eds., *Combinatorial Auctions*, chapter 20, 507–538 (MIT Press, Cambridge, MA).

Churchill AM, Lovell DJ, Mukherjee A, Ball MO (2012) Determining the number of airport arrival slots. *Transportation Science* 47(4):526–541.

Cramton P, Shoham Y, Steinberg R (2006) *Combinatorial Auctions* (MIT Press, Cambridge, MA).

Ferguson J, Kara AQ, Hoffman K, Sherry L (2013) Estimating domestic US airline cost of delay based on European model. *Transportation Research Part C: Emerging Technologies* 33:311–323.

Jacquillat A, Odoni A (2015) An integrated scheduling and operations approach to airport congestion mitigation. *Operations Research* 63:1390–1410.

Mukherjee A, Lovell DJ, Ball MO, Odoni AR, Zerbib G (2005) Modeling delays and cancellation probabilities to support strategic simulations. *Proceedings of 6th USA/Europe Air Traffic Management R & D Seminar, 2005.*

Rassenti SJ, Smith VL, Bulfin RL (1982) A combinatorial auction mechanism for airport time slot allocation. *The Bell Journal of Economics* 13(2):402–417.

Multimodal Transportation Services

TC1: Airport Operations

Thursday 2:45 – 4:15 PM

Session Chair: Senay Solak

2:45 Efficiency, Equity and On-Time Performance Objectives in Airport Demand Management
[1]Alexandre Jacquillat, [2]Vikrant Vaze*
[1]Carnegie Mellon University-Heinz College, [2]Dartmouth College

3:15 A Data-Splitting Algorithm for Flight Sequencing and Scheduling on Two Runways
Rakesh Prakash, Jitamitra Desai*
Nanyang Technological University

3:45 Lower Cost Departures for Airlines: Optimal Gate and Metering Area Allocation Policies Under Departure Metering Concept
[1]Heng Chen, [2]Senay Solak*
[1]University of Nebraska-Lincoln, [2]University of Massachusetts Amherst

Efficiency, Equity and On-time Performance Objectives in Airport Demand Management

Alexandre Jacquillat[1] and Vikrant Vaze[2]

[1]Carnegie Mellon University, Heinz College
[2]Dartmouth College, Thayer School of Engineering

1 Introduction

Absent opportunities for increases in airport capacity or operating enhancements, demand management can be used to better match airport demand and capacity. It commonly takes the form of scheduling interventions that control peak-hour scheduling levels. Busy airports outside the United States operate under slot control policies. In contrast, airline schedules are weakly constrained at US airports. A few airports use "flight caps", but they are much less severe than at slot-controlled airports. As a result of these regulatory differences, US airports achieve higher capacity utilization, but also face larger delays than their European counterparts (Morisset and Odoni, 2011).

Recent research has showed that performance improvements could be achieved at the busiest US airports through limited scheduling interventions that involve only temporal shifts in demand (i.e., changes in the intra-day timetabling of flights), and no reduction in overall demand (i.e., no change in the set of flights scheduled in the day) (Vaze and Barnhart, 2012; Swaroop et al., 2012; Pyrgiotis and Odoni, 2016; Jacquillat and Odoni, 2015). However, existing approaches are focused exclusively on overall scheduling levels at the airports, without considering explicitly the impact of the interventions on the different airlines. In turn, they may penalize one airline (or a small number of airlines) disproportionately.

This paper develops and solves a set of optimization models that incorporate inter-airline equity considerations into airport scheduling interventions. We consider a scheduling process that starts with the preferred schedule of flights, provided by the airlines, and that proposes some scheduling adjustments to reduce anticipated delays. We develop performance indicators based on efficiency, inter-airline equity, and on-time performance, and propose an original lexicographic modeling architecture to optimize scheduling interventions across these objectives. We then perform a theoretical analysis to show that, under some scheduling conditions, efficiency and equity can be jointly maximized and, conversely, that a trade-off between these two objectives can arise in under certain conditions. Last, we generate and solve real-world full scale computational scenarios at New York City's John F. Kennedy Airport (JFK) and show that, under a wide range of realistic and hypothetical scheduling conditions, the consideration of efficiency-based objectives exclusively in airport scheduling interventions may lead to highly inequitable outcomes, but that inter-airline equity can be achieved at no (or minimal) efficiency losses using our models. This suggests that existing approaches for scheduling interventions can be effectively extended to include inter-airline equity considerations.

2 Model Development

The starting point of our model is the *Integrated Capacity Utilization and Scheduling Model (ICUSM)* developed in previous research by Jacquillat and Odoni (2015). It provides a modeling framework for optimizing congestion-mitigating scheduling interventions, but does not account for inter-airline equity considerations.

To address this limitation, we first define the following inputs and variables:

$\mathcal{F}$	=	set of flights, indexed by $i = 1, ..., F$
$\mathcal{A}$	=	set of airlines, indexed by $a = 1, ..., A$
$\mathcal{F}_a$	=	set of flights scheduled by airline a at the airport under consideration
v_i	=	valuation of flight i
u_i	=	displacement (positive or negative) of flight i, as number of 15-minute periods

We propose a set of three performance attributes for scheduling interventions: (i) efficiency (i.e., meeting airline scheduling preferences), (ii) inter-airline equity (i.e., balancing scheduling adjustments fairly among the airlines), and (iii) on-time performance (i.e., mitigating airport congestion). We characterize the trade space between these three attributes by developing a lexicographic optimization approach that:

1. Fixes on-time performance targets: We impose peak expected arrival and departure queue length targets, denoted by A_{MAX} and A_{MAX}, respectively.

2. Maximizes efficiency, subject to scheduling constraints, network connectivity constraints, and on-time performance targets: We formulate the efficiency-maximizing problem by lexicographically maximizing, first, the largest displacement that any flight will sustain, and, second, the total weighted displacement of the schedule. We denote by δ^* and Δ^* their optimal values, respectively.

$$\begin{aligned}
\min \quad & \max_{i \in \mathcal{F}} |u_i| \\
\text{s.t.} \quad & \text{Scheduling and network connectivity constraints} \\
& \text{On-time performance constraints}
\end{aligned}$$

$$\begin{aligned}
\min \quad & \sum_{i \in \mathcal{F}} v_i \, |u_i| \\
\text{s.t.} \quad & \text{Scheduling and network connectivity constraints} \\
& \text{On-time performance constraints} \\
& \text{Min-max efficiency objectives: } |u_i| \leq \delta^*, \forall i \in \mathcal{F}
\end{aligned}$$

3. Maximizes inter-airline equity, subject to scheduling constraints, network connectivity constraints, on-time performance constraints, and efficiency targets. We lexicographically minimize the disutilities borne by the airlines, denoted by σ_a and quantified as the weighted per-flight displacement, i.e., $\sigma_a = \frac{1}{|\mathcal{F}_a|} \sum_{i \in \mathcal{F}_a} v_i \, |u_i|$ for all $a \in \mathcal{A}$. To characterize the trade space between efficiency and equity, we impose that min-max efficiency must be optimal and we denote by $\rho \in [0, \infty)$ the relative loss in weighted efficiency that is allowed. When $\rho = \infty$, we only maximize equity (without any weighted efficiency consideration). When $\rho = 0$, we

maximize equity under optimal efficiency.

$$\begin{aligned}
\operatorname{lex\,min} \quad & \left(\frac{1}{|\mathcal{F}_a|}\sum_{i\in\mathcal{F}_a} v_i\,|u_i|\right)_{a\in\mathcal{A}} \\
\text{s.t.} \quad & \text{Scheduling and network connectivity constraints} \\
& \text{On-time performance constraints} \\
& \text{Min-max efficiency objectives: } |u_i| \leq \delta^*, \forall i \in \mathcal{F} \\
& \text{Weighted efficiency objectives: } \sum_{i\in\mathcal{F}} v_i\,|u_i| \leq (1+\rho)\,\Delta^*
\end{aligned}$$

3 Theoretical Results

First, we show that efficiency and equity can be jointly maximized in the absence of network connections, under uniform flight valuations, and some scheduling conditions detailed in the following two propositions. Figure 1 provides an illustration of the scheduling conditions considered in each of these two propositions. We also introduce the following notations. We denote the number of flights scheduled in period t by $\mathcal{D}_t$ and the capacity limit in period t by $\widehat{\lambda}_t$.

Proposition 3.1 (Figure 1a) shows that efficiency and equity can be jointly maximized if the number of flights scheduled over every set of three consecutive time periods is lower than the total number of flights that can be scheduled over the same three periods. In that case, the imbalances between demand and capacity are small enough so no time period is such that some flights get displaced *to* that period and some other flights get displaced *from* that period.

Proposition 3.1 *If $\sum_{l=t-1}^{t+1} |\mathcal{D}_l| \leq \sum_{l=t-1}^{t+1} \widehat{\lambda}_l, \forall t \in \mathcal{T}$, then there exists a solution that simultaneously solves the efficiency-maximizing and the equity-maximizing problems.*

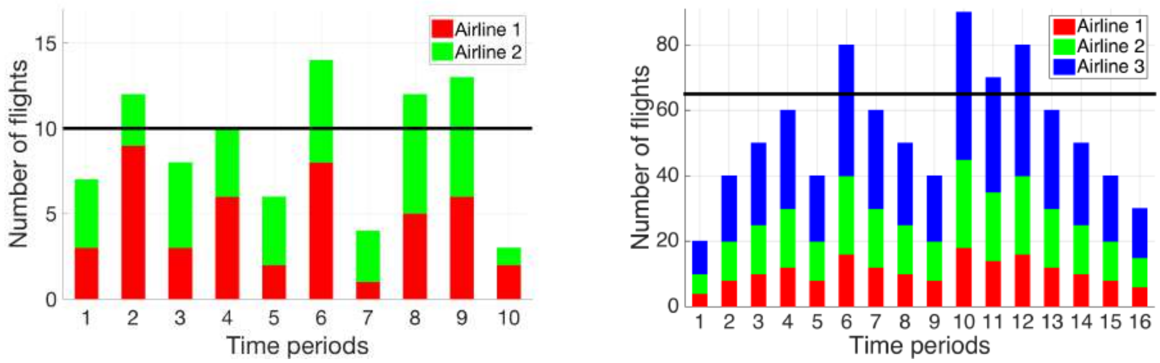
Proposition 3.2 (Figure 1b) shows that efficiency and equity can be jointly maximized if the flight distribution across periods is the same for all airlines. In that case, the schedules of flights of the airlines exhibit the same intra-day variations, which provides significant flexibility in terms of choosing the airlines whose flights should be rescheduled.

Proposition 3.2 *If $\delta^* = 1$ period and there exist integers $(\alpha_a)_{a\in\mathcal{A}}$ and $(\beta_t)_{t\in\mathcal{T}}$ such that $|\mathcal{D}_t \cap \mathcal{F}_a| = \alpha_a\beta_t, \forall a \in \mathcal{A}, t \in \mathcal{T}$, then there exists a solution that simultaneously solves the efficiency-maximizing and the equity-maximizing problems.*

Conversely, we also show that a trade-off between efficiency and equity can arise through (i) inter-airline differences in intra-day flight schedule variations, (ii) network connections, and (iii) intra-airline variations in flight valuations.

4 Computational Results

We then generate computational scenarios at JFK to analyze the performance of the proposed mechanism. We use flight-level data from the Aviation System Performance Metrics (ASPM) (Federal Aviation Administration, 2013) and passenger-level data from the database developed by Barnhart et al. (2014).



(a) Proposition 3.1 (b) Proposition 3.2

Figure 1: Example of scheduling conditions of Propositions 3.1 and 3.2

We first consider the case where all flights are equally valued (i.e., if $v_i = 1, \forall i \in \mathcal{F}$), which corresponds to the current scheduling environment based on the "a flight is a flight" paradigm. Table 1 reports three sets of results, for different values of the on-time performance targets A_{MAX} and D_{MAX}: (i) a solution that *minimizes* inter-airline equity under optimal efficiency (Problem $\overline{\mathbf{P2}}$), (ii) a solution obtained under efficiency objectives alone (Problem $\mathbf{P2}$), and (iii) a solution that maximizes inter-airline equity under optimal efficiency (Problem $\widehat{\mathbf{P3}(\rho^*)}$)). The table reports the total schedule displacement faced by each airline (that is, $\Delta = \sum_{i\in\mathcal{F}} |u_i|$), each airline's disutility (i.e., its weighted per-flight displacement σ_a), and the ratio of the largest to smallest disutility across airlines. These results show that, for any set of values of A_{MAX} and D_{MAX} that we considered, the modeling approach developed in this paper provides strong equity gains at no loss in efficiency. Specifically: (i) the set of efficiency-maximizing solutions includes highly inequitable outcomes, reflected by large values of the max-min ratio $\frac{\max_a \sigma_a}{\min_a \sigma_a}$ under Problem $\overline{\mathbf{P2}}$, (ii) considering efficiency objectives only (Problem $\mathbf{P2}$) does not result in the most inequitable outcome in that set, but can still lead to low inter-airline equity, and (iii) the equity-maximizing solution (Problem $\widehat{\mathbf{P3}(\rho^*)}$) results in high inter-airline equity (e.g., low values of the ratio $\frac{\max_a \sigma_a}{\min_a \sigma_a}$) and the same total displacement as the efficiency-maximizing solution (Problem $\mathbf{P2}$) in all cases considered. Therefore, joint optimization of efficiency and equity is achievable under current schedules of flights and uniform flight valuations (which is the assumption widely used in current practice). In other words, even though the conditions of Propositions 3.1 and 3.2 are not verified, the insights of joint maximization of efficiency and equity still hold in realistic conditions.

Second, we consider the case where all flights are not equally valued. This captures potential extensions of existing and other previously proposed mechanisms for airport scheduling interventions that would allow the airlines to provide the relative timetabling flexibility of their flights (e.g., through an auction-based or credit-based mechanism). In contrast to the previous case of uniform flight valuations, efficiency and inter-airline equity can no longer be jointly maximized, but results (not reported in this abstract) show that accounting for inter-airline equity can significantly improve the outcome of scheduling interventions. Specifically, we find that ignoring inter-airline equity (i.e., optimizing efficiency-based objectives alone) may lead to highly inequitable outcomes, and that inter-airline equity can be achieved at comparatively small losses in efficiency.

Table 1: Number of flights displaced and airline disutilities per airline under uniform valuations

On-time targets			Number of flights displaced					Disutility: $\sigma_a = \frac{1}{\lvert\mathcal{F}_a\rvert}\sum_{i\in\mathcal{F}_a}\lvert u_i\rvert$				
A_{MAX}	D_{MAX}	Model	DAL	AAL	JBU	Others	All	DAL	AAL	JBU	Others	$\frac{\max_a \sigma_a}{\min_a \sigma_a}$
14	23	$\overline{\mathbf{P2}}$	1	13	1	5	20	0.3%	5.0%	0.6%	1.2%	16.00
		$\mathbf{P2}$	1	9	2	8	20	0.3%	3.5%	1.1%	2.0%	11.08
		$\widehat{\mathbf{P3}(\rho^*)}$	4	5	3	8	20	1.3%	1.9%	1.7%	2.0%	1.57
13	20	$\overline{\mathbf{P2}}$	1	29	9	7	46	0.3%	11.2%	5.2%	1.7%	35.69
		$\mathbf{P2}$	7	18	8	13	46	2.2%	6.9%	4.6%	3.2%	3.16
		$\widehat{\mathbf{P3}(\rho^*)}$	13	10	7	16	46	4.1%	3.8%	4.0%	3.9%	1.06
12	18	$\overline{\mathbf{P2}}$	1	28	27	9	65	0.3%	10.8%	15.5%	2.2%	49.66
		$\mathbf{P2}$	10	27	10	18	65	3.1%	10.4%	5.7%	4.4%	3.32
		$\widehat{\mathbf{P3}(\rho^*)}$	18	14	10	23	65	5.6%	5.4%	5.7%	5.6%	1.07
11	15	$\overline{\mathbf{P2}}$	37	113	39	17	206	11.6%	43.5%	22.4%	4.2%	10.43
		$\mathbf{P2}$	50	57	32	67	206	15.6%	21.9%	18.4%	16.4%	1.40
		$\widehat{\mathbf{P3}(\rho^*)}$	57	46	31	72	206	17.8%	17.7%	17.8%	17.6%	1.01

In turn, the approach developed in this paper offers the potential to extend existing approaches to airport demand management (either the slot control policies in place at busy airports outside the United States, or the scheduling practices at a few of the busiest US airports where flight caps are in place) in a way that balances scheduling interventions fairly among the airlines, thus considerably enhancing their applicability in practice.

References

Barnhart, C., Fearing, D., and Vaze, V. (2014). Modeling Passenger Travel and Delays in the National Air Transportation System. *Operations Research*, 62(3):580–601.

Federal Aviation Administration (2013). Aviation System Performance Metrics (ASPM) database. Accessed April 4, 2013. Available at: https://aspm.faa.gov/apm/sys/main.asp.

Jacquillat, A. and Odoni, A. (2015). An Integrated Scheduling and Operations Approach to Airport Congestion Mitigation. *Operations Research*, 63(6):1390–1410.

Morisset, T. and Odoni, A. (2011). Capacity Delays and Schedule Reliability at Major Airports in Europe and the United States. *Transportation Research Record*, 2214:85–93.

Pyrgiotis, N. and Odoni, A. (2016). On the Impact of Scheduling Limits: A Case Study at Newark International Airport. *Transportation Science*, 50(1):150–165.

Swaroop, P., Zou, B., Ball, M., and Hansen, M. (2012). Do More US Airports Need Slot Controls? A Welfare Based Approach to Determine Slot Levels. *Transportation Research Part B: Methodological*, 46(9):1239–1259.

Vaze, V. and Barnhart, C. (2012). Modeling Airline Frequency Competition for Airport Congestion Mitigation. *Transportation Science*, 46(4):512–535.

A Data-Splitting Algorithm for Flight Sequencing and Scheduling on Two Runways

Rakesh Prakash and Jitamitra Desai
School of Mechanical and Aerospace Engineering
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
*Corresponding Author: jdesai@ntu.edu.sg

1. Introduction

It is widely recognized that rising air traffic demand has placed a significant stress on the entire *Air Traffic Management* (**ATM**) system, costing airlines, passengers, and the overall economy several billions of dollars each year. While injecting additional capacity into the ATM system through infrastructure development can relieve the stress placed on the system, it has become fairly evident that there is still sufficient scope for improvement within the existing ATM system by squeezing additional capacity from critical bottleneck operations related to arrivals, departures, runways, and taxiways. One such aspect is optimizing sequencing and scheduling of flight operations on runways, which is commonly referred to in the literature as the *Aircraft Sequencing Problem* (**ASP**). Specifically, in the *static version* of this problem; given a set of aircraft, along with information on the earliest/latest operation time for each aircraft (be it an arrival or a departure), and the minimum safety regulations to protect trailing aircraft from wake vortices generated by leading aircraft; the objective is to determine a sequence (that optimizes a predefined objective) while simultaneously achieving safety, efficiency, and equity in the ATM system. Passenger safety is achieved by maintaining the required separations between aircraft; runway efficiency is equivalent to achieving low average delay or high throughput; and airline equity is modeled by implementing a *constrained position shifting* (CPS) strategy wherein an aircraft cannot be shifted by more than k positions (the so-called *maximum position shifting* (MPS) parameter) from its initial FCFS-based position.

Many variations of the ASP can be postulated depending upon the number of runways (single or multiple); mode of runway operations (segregated or mixed); problem objectives (minimizing delay or maximizing throughput); and constraints such as inclusion of time windows, permissibility of early landings, and CPS requirements. While the ASP on a single runway has been extensively studied in the literature ([1–4]), the multiple runway ASP (***m*-ASP**) has received scant attention, and the works that exist only consider the simpler objective function of minimizing weighted delay without imposing any CPS requirements ([5, 6]). Note that the m-ASP is a theoretically harder problem to solve as compared to the 1-ASP (single runway ASP) because of the additional runway allocation decisions ([7]) that need to be incorporated into the model formulation. The equivalence of the m-ASP with the well-known vehicle routing problem with time-windows (VRPTW) has already been established, which renders this an NP-Hard problem. The m-ASP is also a more practically relevant problem than its single runway variant because most of the busy international airports have at least two runways.

Keeping in view the importance as well as the challenges posed in solving the generic m-ASP, in this paper, we consider a special case of the m-ASP on two runways (which we refer to as 2-ASP), while accounting for both segregated and mixed-mode of operations, CPS constraints, wide time-windows, early landings/departures, and the objective of maximizing total throughput from both runways. To the best of our knowledge, no efficient solution exists for solving the 2-ASP under the aforementioned scenarios. We begin by formulating our problem as a 0-1 *mixed-integer program* (MIP), which is an adaptation of the model in [2] enhanced with the inclusion of the CPS constraints and symmetry breaking constraints. Recognizing that solving this model is not computationally viable for large-scale instances, the main thrust of this paper is an extension of the massively parallelizable data-splitting algorithm (DS-ASP), originally proposed for the 1-ASP, that optimizes flight sequences by a repeated application of this 0-1 MIP on smaller data sets, while accounting for the direct and induced effects of these smaller data sets on one another, and demonstrates the efficacy of this approach.

The remainder of this paper is organized as follows. In Section 2, we present the 0-1 MIP formulation for the 2-ASP along with some preliminary computations. Next, in Section 3, the details of the data-splitting algorithm, and its pseudo-code are described. Then, Section 4 demonstrates some computational results for large-scale (realistic) instances, and finally, Section 5 summarizes the contributions of this work and suggests extensions for future research.

2. Optimization Problem

We are now ready to formulate the 2-ASP as a 0-1 mixed integer program, as detailed below.

Description of Index Sets and Parameters

- $\mathcal{F}$: Set of all arriving and departing flights
- $\mathcal{R}$: Set of all runways
- E_i : Earliest time of arrival (departure) of aircraft i
- T_i : Target time of arrival (departure) of aircraft i
- L_i : Latest time of arrival (departure) of aircraft i
- Δs_{ij} : Required safety separation (in seconds) at runway threshold, if flight i is ahead of flight j
- seq_i : Position of flight i based on the FCFS sequence
- k : Specified maximum position shifting (MPS) parameter

Decision Variables

$$\bullet\ x_{ij} = \begin{cases} 1, & \text{if flight } i \text{ is ahead of flight } j \text{ in sequence} \\ 0, & \text{otherwise} \end{cases} \tag{1a}$$

$$\bullet\ y_{ir} = \begin{cases} 1, & \text{if flight } i \text{ lands on runway } r \\ 0, & \text{otherwise} \end{cases} \tag{1b}$$

$$\bullet\ w_{ij} = \begin{cases} 1, & \text{if flight } i \text{ and flight } j \text{ land on same runway} \\ 0, & \text{otherwise} \end{cases} \tag{1c}$$

- t_i = Scheduled time of arrival (departure) of flight i (1d)
- z_r = Makespan of runway r (1e)

2-ASP: Minimize

$$z_1 \tag{2a}$$

subject to:

$$x_{ij} + x_{ji} = 1, \quad \forall\, i < j,\ (i,j) \in \mathcal{F} \tag{2b}$$

$$\sum_{r \in \mathcal{R}} y_{ir} = 1, \quad \forall\, i \in \mathcal{F} \tag{2c}$$

$$w_{ij} \geq y_{ir} + y_{jr} - 1, \quad \forall\, i < j,\ (i,j) \in \mathcal{F},\ \forall\, r \in \mathcal{R} \tag{2d}$$

$$w_{ij} \leq y_{ir} - y_{jr} + 1, \quad \forall\, i < j,\ (i,j) \in \mathcal{F},\ \forall\, r \in \mathcal{R} \tag{2e}$$

$$w_{ij} = w_{ji}, \quad \forall\, i < j,\ (i,j) \in \mathcal{F} \tag{2f}$$

$$t_j \geq t_i + \Delta s_{ij} w_{ij} - M_1(1 - x_{ij}), \quad \forall\, (i,j) \in \mathcal{F} \tag{2g}$$

$$z_r \geq t_i - M_2(1 - y_{ir}), \quad \forall\, i \in \mathcal{F},\ \forall\, r \in \mathcal{R} \tag{2h}$$

$$z_1 \geq z_2 \tag{2i}$$

$$-k \leq (n - \sum_{j \in \mathcal{F},\, i \neq j} x_{ij}) - seq_i \leq k, \quad \forall\, i \in \mathcal{F} \tag{2j}$$

$$E_i \leq t_i \leq L_i, \quad \forall\, i \in \mathcal{F} \tag{2k}$$

$$x_{ij}, w_{ij} \in \{0,1\},\ \forall\, (i,j) \in \mathcal{F};\ y_{ir} \in \{0,1\},\ z_r,\ t_i \geq 0,\ \forall\, i \in \mathcal{F},\ r \in \mathcal{R}. \tag{2l}$$

In the above formulation, the objective function (2a) seeks to minimize the makespan; constraint (2b) enforces the order precedence relationship between flights i and j; constraint (2c) dictates that each plane is assigned to exactly one runway; constraints (2d) and (2e) together ensure that the conditional relationships between the w- and y- variables are respected; constraint (2f) is a symmetry constraint that maintains consistency of runway allocations; constraint (2g) mandates the time-based separation requirements between flights at the runway (as specified by the FAA for various flight classes), where $M_1 \equiv (L_i + \Delta s_{ij} - E_j)$, are satisfied; constraint (2h) defines the makespan of runway r to be greater than or equal to the scheduled times of all flights allocated to runway r, where $M_2 \equiv L_i$; constraint (2i) breaks the symmetry by specifying that the last scheduled flight in the overall schedule must be assigned to the first runway; constraint (2j) imposes the CPS constraint that an aircraft cannot be shifted by more than k positions from its initial (FCFS) position, where $n \equiv |\mathcal{F}|$; constraint (2k) maintains the scheduled time of arrival (departure) at the runway to be between the earliest and latest times for each aircraft; and finally constraint (2l) imposes binary and non-negativity restrictions on the $(w,\ x,\ y)$- and $(t,\ z)$- variables, respectively.

3. Algorithm

There are two stages involved in the proposed data-splitting algorithm (**DS-ASP**). The first stage is *instance generation*, in which the original flight data-set is divided into several possible pairs of disjoint subsets of leading and following aircraft. The second is the *solution stage*, in which each resulting subset pair is independently solved to determine its optimal solution. The obtained solutions are finally compared to get the overall optimum. The underlying drivers for using this algorithm are: (i) Given two sets of aircraft, under the condition that one set has to necessarily follow the other, determining the optimal solution for each set separately, while accounting for the effect of the resulting schedule on their respective objective functions, is always computationally more efficient; (ii) As a result of the CPS constraint, the number of pairs of leading and following data-sets are very limited, i.e., of the order of $\binom{2k}{k}$, where k is value of the MPS parameter, and is *independent* of the total number of flights n; (iii) The 0-1 MIP formulation (Problem 2-ASP) is computationally very efficient when solving smaller-sized instances ($\leq$ 15 aircraft); and (iv) It is massively parallelizable.

3.1 Algorithmic outline for arriving traffic

Stage A: Preparation of instance pairs

Let n denote the number of aircraft positioned at $\{1,...,m-1,m,m+1,...,n\}$ in FCFS order. After splitting this flight data into two (possibly unequal) halves, the leading set A and following set B are composed of aircraft at positions $\{1,...,m\}$ and $\{m+1,...,n\}$, respectively. Owing to the CPS constraint, as only aircraft between $(m-k+1)$ and $(m+k)$ positions can crossover between the leading/following sets, there are only $\binom{2k}{k}$ such pairs. Note that both the *aircraft type* and the *arrival time* of the last scheduled flight on each runway in the leading set affect the makespan of the following set (as well as the overall makespan) because aircraft in the following set have to maintain the required safety separations from their counterparts in the leading set. Hence, in order to account for such phenomenon, the number of pairs are further enlarged by fixing all possible aircraft in the leading set that can be positioned last on each runway, and a constraint is added to the optimization problem corresponding to the following set to ensure that all the flights arrive after the landing time of the last aircraft in the leading set. Specifically, observing that only aircraft between positions $(m-k)$ to m can occupy the last position in set A, each pair (A, B) results in additional combinations, denoted as ($\dot{A}$, B). Furthermore, each pair ($\dot{A}$, B) gives rise to further combinations ($\ddot{A}$, B) after fixing each aircraft type which can land at the last position on the second runway. These ($\ddot{A}$, B) pairs are what will finally be used by the DS-ASP algorithm. This pairing scheme is illustrated in the example below.

Consider an instance of eight aircraft $\{H1,H2,H3,L4,L5,H6,S7,S8\}$, where each aircraft is represented by its aircraft type (H: Heavy, L: Large, S: Small) and its relative position in the FCFS sequence. Assuming $k=2$, there are $\binom{4}{2}=6$ resulting pairs (A, B), which are given by:

(**i**) $A=\{H1,H2,H3,L4\},\ B=\{L5,H6,S7,S8\}$; (**ii**) $A=\{H1,H2,H3,L5\},\ B=\{L4,H6,S7,S8\}$; (**iii**) $A=\{H1,H2,H3,H6\},\ B=\{L4,L5,S7,S8\}$; (**iv**) $A=\{H1,H2,L4,L5\},\ B=\{H3,H6,S7,S8\}$; (**v**) $A=\{H1,H2,L4,H6\},\ B=\{H3,L5,S7,S8\}$; (**vi**) $A=\{H1,H2,L5,H6\},\ B=\{H3,L4,S7,S8\}$.

For each of these listed pairs (A, B), permuting flights within set A, predicated on the aircraft occupying the last position on the first runway (denoted as $\bullet$), results in several additional combinations, which we denote as ($\dot{A}$, B). As an example, for the case: $A=\{H1,H2,H3,L4\}$ and $B=\{L5,H6,S7,S8\}$, such permutations within set A result in the following three pairs:

$$\dot{A}=\{H1,H2,H3,\dot{L}4\},\ B=\{L5,H6,S7,S8\},$$
$$\dot{A}=\{H1,H2,L4,\dot{H}3\},\ B=\{L5,H6,S7,S8\},$$
$$\dot{A}=\{H1,H3,L4,\dot{H}2\},\ B=\{L5,H6,S7,S8\}.$$

Note that, when $A=\{H1,H2,H3,H6\}$ and $B=\{L4,L5,S7,S8\}$, as aircraft $H6$ cannot be scheduled before position 4 in the sequence (as this would violate the CPS constraint), no further pairings are feasible in this case. Furthermore, if all the aircraft satisfy the ELW rule ([1]), then some of the pairs, e.g., $\dot{A}=\{H1,H3,L4,H2\}$ and $B=\{L5,H6,S7,S8\}$, can be discarded. Now, as aforementioned, for each of these pairs ($\dot{A}$, B), several additional derivative pairs ($\ddot{A}$, B) are generated based on the aircraft type which can be scheduled last on the second runway. As an example, for the pair: $\dot{A}=\{H1,H2,L4,\dot{H}3\}$, $B=\{L5,H6,S7,S8\}$, the resulting derivative ($\ddot{A}$, B) pairs are listed as follows:

$$\ddot{A}=\{H1,H2,\ddot{L}4,\dot{H}3\},\ B=\{L5,H6,S7,S8\},$$
$$\ddot{A}=\{\ddot{H}1,\ddot{H}2,L4,\dot{H}3\},\ B=\{L5,H6,S7,S8\}.$$

Note that in the overall schedule, the landing order of the last scheduled aircraft on the second runway cannot be fixed in advance, and hence more than one aircraft (denoted as $\ddot{\bullet}$) in $\ddot{A}$ resulting from the same aircraft type are potential candidates to be scheduled last on the second runway.

Stage B: Solving the instance pairs

Let ($\ddot{z}_1$, $\ddot{z}_2$) denote an *optimal state* for a leading set $\ddot{A}$, where $\ddot{z}_r$ denotes the optimal makespan on runway r ($r=1,2$), with $\ddot{z}_1 \geq \ddot{z}_2$. When optimizing the leading set of flights, it is always desirable to keep the makespan on the second runway as small as possible so as to accommodate a feasible solution for the following set of flights. However, there is an inverse relationship between $\ddot{z}_1$ and $\ddot{z}_2$, i.e., decreasing one may increase the other, and as a result, a brute-force method that evaluates all possible optimal states ($\ddot{z}_1$, $\ddot{z}_2$) remains the only avenue to find the optimal solution. But, the number of such states can be very large and hence evaluating all possibilities is a computationally intensive task. However, as flights in the following set are constrained to land at or after $\ddot{z}_1$, the key to deciding optimality of the following set is not only to lower the $\ddot{z}_1$-value but rather the gap $\ddot{z}_1-\ddot{z}_2$, because aircraft in the following set that land on the second runway can take advantage of this gap to land as early as possible. This observation lays the foundation of our algorithmic design, wherein a nonnegative continuous indicator variable δ, which reflects the to-be-obtained advantage due to the $\ddot{z}_1-\ddot{z}_2$ gap, is embedded in the optimization runs of the following set. This δ-variable may take on a non-zero value, which indicates that adding δ to the existing gap $\ddot{z}_1-\ddot{z}_2$, i.e. optimizing with respect to the state ($\ddot{z}_1$, $\ddot{z}_2-\delta$) may result in a better optimal solution, whereas a zero value of δ acts as a stopping criteria indicating that no further states resulting from that pair ($\ddot{A}$, B) need to be evaluated. Furthermore, whenever δ is found to be non-zero, the leading set is re-optimized to seek a new improved state ($\ddot{z}_1,\ddot{z}_2$). A formal statement of the pseudo-code of the proposed data-splitting algorithm is given below.

Algorithm 1 Pseudo-code for the proposed data-splitting algorithm.

1: Set iteration counter $p \leftarrow 0$, $OptVal^{(p)} \leftarrow +\infty$, incumbent $\leftarrow OptVal^{(p)}$, S = {set of all pairs $(\ddot{A},B)$}. Go to Step 2.

2: If $S=\emptyset$, stop; return incumbent as best solution. Else, set $p \leftarrow p+1$, and go to Step 3.

3: Arbitrarily select one instance pair $(\ddot{A}_p,B_p)$ from S. Set temp.best, $\ddot{z}_2 \leftarrow +\infty$. Go to Step 4.

4: Solve 2-ASP for set $\ddot{A}_p$ with $OptVal^{(p)}_{\ddot{A}_p}$ = min $\left\{\sum_{i\in\ddot{A}_p}\theta z_1+z_2 : \text{(1a)-(21)},\ z_2\leq\ddot{z}_2-\varepsilon\right\}$, where θ is a (large) weighting parameter and $\varepsilon>0$ is a predefined tolerance. Let ($\ddot{z}_1$, $\ddot{z}_2$) be the optimal solution (state) and denote (l_1,l_2) to indicate the last aircraft on the first and second runways. Go to Step 5.

5: Solve Problem 2-ASP for set B_p with

$$OptVal^{(p)}_{B_p} = \min\left\{\sum_{i\in B_p} z_1+\varepsilon\delta : \text{(1a)-(21)},\ t_i\geq\ddot{z}_1,\ t_i\geq\ddot{z}_1+y_{i1}\Delta s_{l_1 i},\ t_i\geq\ddot{z}_2+y_{i2}\Delta s_{l_2 i}-\delta, \forall i\in B_p\right\},$$

where ε is a preset tolerance. If $\delta>0$ go to Step 6, else go to Step 7.

6: Rerun the optimization problem in Step 5 with constraint $\delta=0$. If $OptVal^{(p)}_{B_p}<$ temp.best, set temp.best $\leftarrow OptVal^{(p)}_{B_p}$. Go to Step 4.

7: If $OptVal^{(p)}_{B_p}<$ temp.best, set temp.best $\leftarrow OptVal^{(p)}_{B_p}$; $OptVal^{(p)}\leftarrow$ temp.best. Go to Step 8.

8: If $OptVal^{(p)}<$ incumbent, set incumbent $\leftarrow OptVal^{(p)}$. Update $S\leftarrow S\setminus\{\ddot{A}_p,B_p\}$. Go to Step 2.

4. Computational Results

In our prototype implementation, we tested the proposed data-splitting algorithm on various randomly generated instances comprising thirty and thirty-five aircraft, under different arrival traffic scenarios. The traffic mixture for these instances was set as: 40% Heavy + 40% Large + 20% Small aircraft, and the target times of aircraft were also randomly generated, assuming that each aircraft appears every γ seconds, where $\gamma = 30$ or 35 seconds. We also assume that a flight can arrive or depart up to 60 seconds earlier and no more than 1800 seconds later than its scheduled target time. All of our computations are performed on a Windows machine, equipped with a Intel Xeon CPU E5-1630v3 3.70GHz 32GB RAM processor, using MATLAB R2011b in conjunction with GuRoBi 6.0.0 as the underlying MIP-solver.

From the results recorded in Table 1, on average, we can observe more than an 81% reduction in the computational time required to solve all randomly generated instances by using the proposed DS-ASP algorithm as compared to directly solving the original 2-ASP formulation (2a)-(2l) using the commercially available solver GuRoBi [8]. Moreover, this computational reduction is in excess of 97%, when the instance pairs are solved in parallel.

Table 1: Computational time comparison between the original MIP model and data splitting algorithm

Mode of Operation		**Without Parallelization**						**With Parallelization**		
	$k\rightarrow$	**1**		**2**		**3**		**1**	**2**	**3**
	n / **Time (s)**↘	**2-ASP**	**DS-ASP**	**2-ASP**	**DS-ASP**	**2-ASP**	**DS-ASP**	**DS-ASP**	**DS-ASP**	**DS-ASP**
Arrival	30	>3600	140.9	>3600	276.9	>3600	1173.5	35.2	47.4	168.3
Arrival	30	>3600	130.2	>3600	428.6	>3600	2081.7	31	81.7	225.5
Arrival	30	>3600	58.9	>3600	216.2	>3600	384	39.6	57.1	59.6
Arrival	30	>3600	74.9	>3600	463.8	>3600	1688.9	25.4	83.5	241.5
Arrival	35	>3600	158.2	>3600	375	>3600	1138.4	35	69	183.9
Arrival	35	>3600	180	>3600	808.4	>3600	2531	65.3	233.6	255.6

5. Conclusions

In this paper, we proposed an extension of the data-splitting algorithm, originally presented in [1], which provides real-time optimal solutions for the aircraft sequencing problem on two runways with the objective of maximizing throughput under both segregated and mixed-traffic conditions. The performance of the DS-ASP algorithm clearly demonstrates the algorithmic speed-up that can be obtained, while achieving optimality in all practical test-bed instances. We are also in the process of further improving the efficiency of the algorithm, notably by reducing the number of instance-pairs and optimal states through model-enhancements, pruning techniques, and associated flight data analysis. Finally, the computational efficiency of this algorithm can be further improved by splitting the data into more than two subsets, and the results of these investigations are forthcoming.

References

[1] J Desai and R Prakash. Flight sequencing and scheduling: A data-driven approach. In *Proceedings of the Industrial and Systems Engineering Research Conference*. Institute of Industrial Engineers, Anaheim, CA, 2016.

[2] John E Beasley, Mohan Krishnamoorthy, Yazid M Sharaiha, and D Abramson. Scheduling aircraft landings- the static case. *Transportation Science*, 34(2):180–197, 2000.

[3] Ahmed Ghoniem, Hanif D Sherali, and Hojong Baik. Enhanced models for a mixed arrival-departure aircraft sequencing problem. *INFORMS Journal on Computing*, 26(3):514–530, 2014.

[4] J Desai and R Prakash. An optimization framework for terminal sequencing and scheduling: The single runway case. In *Proceedings of the Complex Systems Design & Management 2016 Asia Conference*, pages 195–207. Springer, Cham, 2016.

[5] Dirk Briskorn and Raik Stolletz. Aircraft landing problems with aircraft classes. *Journal of Scheduling*, 17(1):31–45, 2014.

[6] Ahmed Ghoniem, Farbod Farhadi, and Mohammad Reihaneh. An accelerated branch-and-price algorithm for multiple-runway aircraft sequencing problems. *European Journal of Operational Research*, 246(1):34–43, 2015.

[7] Harilaos N Psaraftis. A dynamic programming approach for sequencing groups of identical jobs. *Operations Research*, 28(6):1347–1359, 1980.

[8] Z Gu, E Rothberg, and R Bixby. GuRoBi (www.gurobi.com).

Lower Cost Departures for Airlines: Optimal Gate and Metering Area Allocation Policies under Departure Metering Concept

Heng Chen[1], Senay Solak[2]
[1]College of Business Administration, University of Nebraska-Lincoln
[2]Isenberg School of Management, University of Massachusetts Amherst

Extended Abstract

1. Motivation and Problem Description

Air traffic demand is expected to grow twofold over the next two decades (IATA 2014, FAA 2015). Due to this growth, airport surface areas, more specifically runways, have been identified as choke points in airport systems, causing serious congestion at airports. According to Balakrishna et al. (2010), 60% of delay in air traffic is due to airport operations. Also, according to estimates by Idris (2015) and Brinton et al. (2007), taxi-in and taxi-out delays in the U.S. result in extra fuel burn costs on the order of $900 million. These delays have led to significant fuel burn costs for airlines and air pollutant emissions for the public. Hence, there is a great need to improve the efficiency of airport operations to relieve the congestion, reduce fuel consumption and improve customer service.

To reduce delay and improve airport efficiency, the National Aeronautics and Space Administration (NASA) has developed the Airspace Technology Demonstration-2 (ATD-2), aimed at integrating the arrival, departure and surface activities and developing precise schedules for flights at gates, runways, and arrival/departure fixes. *Departure metering*, as a key component of ATD-2, is an airport surface management procedure that limits the number of aircraft on the runway by either holding aircraft at gates or at a predesigned metering area (NATCA 2015). By holding aircraft at gates or at a predesigned metering area with engine idle, the departure metering procedure can reduce fuel burn costs for airlines and airports through shortening runway queues and decreasing unnecessary stops and waits with aircraft engine on. In addition, by integrating the gate, taxiway, and runway activities, the procedure can also improve the coordination and communication between different functions at airports.

Field tests have shown significant fuel benefits and suggested an important role for this procedure in the Next Generation Air Transportation System (NextGen). The six-month long departure metering program at John F. Kennedy International Airport (JFK) has shown to lower fuel burn costs by $10-15 million, and carbon dioxide emissions by 48,000 metric tons. In addition, the program is also expected to result in significant reduction in delays due to reduced taxing hours (Nakahara et al. 2011). Several other airports are also testing departure metering procedures (Lozito 2016).

Our motivating hypothesis in this study is that there is a potentially significant value for airlines in using certain departure policies during a departure metering implementation. A key concern in departure operations is how to allocate aircraft such that efficiency is improved while throughput is being maintained, where efficiency is defined as a function of fuel costs, emissions, noise, and runway utilization. This is a difficult dynamic problem where uncertainties of new arrivals and pushback delay need to be taken into account. In this study, we address this operational problem and identify policies that would enable improved efficiency for airlines and reduced environmental impacts in flight departure operations.

Given these observations, in this paper we seek answers to the following research questions: Given the set of aircraft scheduled to arrive and depart at an airport, *which aircraft should be allowed to push back from the gate, which aircraft should be allocated a gate, and which aircraft should be sent to the runway from the metering area?* Furthermore, from a strategic planning perspective, *what is the optimal metering area capacity to be used in implementing departure metering operations?* We develop a dynamic programming framework to answer these questions. In addition, numerical analyses are also presented to quantify the potential savings that can be achieved through the proposed optimal departure policies.

2. Contributions and Significance of Research

The main contributions of this paper are as follows: (1) This is the first paper that captures the stochasticity in departure metering operations and derives optimal policies to improve efficiency for airlines and the society; (2) Unlike most of the existing studies on departure operations at airports, we address some additional decisions that can create value, such as controlling the departure flow through the use of a metering area; (3) Our study also adds to the limited literature on stochastic modeling of departure operations, as arrivals and pushback delays are captured under a stochastic optimization framework; (4) Several practical departure policies are introduced which can be easily implemented by air traffic controllers without referring to computerized tool or advanced training, and still produce considerable savings compared to current practices.

3. Modeling Approach

Our modeling is based on a finite horizon Markov decision process (MDP) formulation of the problem, for which we obtain both exact and easy-to-implement heuristic solutions. As part of the model description, we first provide the following additional information on the operational framework and the corresponding decision process.

Consider the following decision problem faced by an airport controller. At a given decision epoch, the controller observes the amounts of aircraft scheduled to arrive, waiting at the taxiway to be assigned a gate, at gates, at the metering area, on the runway, and planned to depart, respectively. We define the above information as the distribution of aircraft at the airport, which is consistently changing with the air traffic flow. For a new arrival at the airport, the controller either guide it to stay at the taxiway if there is no available gate at the moment or move it to a gate once there is an empty gate. Note there might be a number of aircraft in a queue waiting for gates when the gates are fully utilized. For the aircraft held at gates which are ready to pushback, the controller has three options: continue staying them at the gate, move them to the metering area, or direct them to join the departure queue directly. For the aircraft at the metering area, the controller can either direct them to the runway or continue staying them at the metering area if there is runway congestion. For the aircraft on the runway, the controller will schedule their departure once there are departure slots available. During this process, the controller can determine the number of aircraft to be pushed back to the metering area from the gates and the number of aircraft to be directed to the runway from the metering area to achieve the desired/target distribution of aircraft for the next decision epoch. Different actions can incur different fuel, environmental and other relevant costs. However, there are several uncertain factors affecting the traffic flow, such as the weather and human factors. In this paper we consider primarily two stochastic issues, the number of actual arrivals and the number of actual pushback aircraft. Factors such as the number of scheduled departures are considered deterministic. Due to the uncertainty involved, the realized distribution of aircraft at the airport for the next decision epoch can be different from the target.

Within the dynamic programming framework used, for modeling purposes, we discretize the time horizon into discrete time periods, each with a fixed duration. We also assume that the controller observes the distribution of aircraft and make corresponding decisions at the beginning of each period. The state set in each stage is defined as the distribution of aircraft at the airport before taking any actions, i.e. the number of aircraft waiting for gates, the number of available gates, the number of aircraft at the metering area, and the number of aircraft on the runway. The action set is defined as the desired or target distribution of aircraft at the airport for the given period to reduce congestion and ensure efficient flow of operations. More specifically, the controllers can make the following two decisions to affect the allocation of aircraft, namely the number of aircraft to be pushed back to the metering area from the gates, and the number of aircraft to be directed to the runway from the metering area.

The transition probabilities are modeled and determined based on the probability distribution for the number of arrivals in a given period for a given arrival rate and the probability distribution for the number of aircraft to pushback in a given period as discussed in Sölveling et al. (2011). The cost structure is based on the costs of holding aircraft at different facilities, which include cost of holding on the taxiway, cost of holding at gates, cost of holding at the metering area and runway holding cost. The overall objective in this MDP representation is to find an optimal mapping of states to target departure metering policies for each decision epoch. The optimal policies can be obtained by solving an optimality equation numerically through backward induction.

While the optimal policies identified through the solution of the optimality equations above provide the lowest cost policies, air traffic controllers may find these policies difficult to implement as they are based on numerical solutions and a computerized tool which is necessary for overall implementation. In this paper we introduce four easy-to-implement departure metering policies as an alternative tool, including MaxiRunway policy, N-Control policy, Low-Cost policy and (s, S) policy. We then implement a comparative analysis between these practical policies and the optimal numerical solutions. We also quantify the potential value created by these policies over current practices.

In addition, we identify the optimal metering area capacity using marginal analysis to minimize expected overall costs from a strategic perspective.

4. Summary of Major Results

In this paper we study optimal departure metering policies at airports from both tactical and strategic perspectives. We develop a stochastic dynamic programming framework to identify such optimal policies, while also studying some near-optimal practical policies for airlines from a tactical perspective. We implement a comparative analysis between four practical policies and the optimal numerical solutions, and find that the (s, S) policy can produce considerable savings compared to current practices. We also look at how the optimal departure metering policies change with respect to different state variables. Furthermore, we introduce an enumeration procedure to identify the optimal capacity for the departure metering area.

Using the developed optimal policies, we perform extensive simulations based on the departure implementation at the Detroit Metropolitan Wayne County Airport (DTW). Our findings show that a capacity of 7 aircraft is the best departure metering configuration at this airport. Savings for airlines due to such policies can be around $30.8 million if these policies are adapted by top ten major airports in the U.S.

Through our analysis, we find that utilization of the proposed optimal policies could add to the value of departure metering procedures and improve overall efficiency by around 14-20% over the current practice as described by Nakahara et al. (2011). Given the need for smooth and integrated surface operations by airlines and airports, the proposed optimal departure metering policies can add to the value of NASA's ATD-2 implementation by improving overall efficiency and sustainability of departure operations.

References

Balakrishna, P., R. Ganesan, L. Sherry. 2010. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures. Transportation Research Part C: Emerging Technologies 18(6) 950–962.

Brinton, C., L. Cook, S. Atkins. 2007. Collaborative airport surface metering for efficiency and environmental benefits. Proceedings of the 2007 Integrated Communications, Navigation and Surveillance Conference.

FAA. 2015. FAA aerospace forecast Fiscal Years 2015-2035. Tech. rep., Federal Aviation Administration.

IATA. 2014. New IATA passenger forecast reveals fast-growing markets of the future. Press Release No.:57, 16 October 2014.

Idris, H. 2015. Identification of local and propagated queuing effects at major airports. AIAA Aviation Forum, Aviation Technology, Integration and Operations (ATIO) conference, Dallas, TX.

Lozito, S. 2016. NASA Ames Aviation Systems Division: ATD-2 main. URL http://www.aviationsystemsdivision.arc.nasa.gov/research/tactical/atd2.shtml. Retrieved March 20, 2016.

Nakahara, A., T. Reynolds, T. White, C. Maccarone, R. Dunsky. 2011. Analysis of a surface congestion management technique at New York JFK Airport. Proceedings of the 11th AIAA Aviation Technology, Integration and Operations (ATIO) Conference.

NATCA. 2015. NATCA Bookshelf - National office week in review: Oct. 21, 2015. URL http://natca.uberflip.com/i/589908-national-office-week-in-review-oct-21-2015/19. Retrieved March 20, 2016.

Sölveling, G., S. Solak, J-P. Clarke, E. Johnson. 2011. Runway operations optimization in the presence of uncertainties. Journal of Guidance, Control, and Dynamics 34(5) 1373–1382.

Multimodal Transportation Services

TD1: Bus and Taxi Transport

Thursday 4:30 – 6:00 PM

Session Chair: Mark Hickman

4:30 Choice between Metro and Taxi under Travel Time Variability
Gege Jiang, Hong Lo*
Hong Kong University of Science and Technology

5:00 Multi-Cycle Optimal Taxi Routing with E-Hailing
[1]Xinlian Yu, [1]Song Gao, [1]Hyoshin Park, [2]Xianbiao Hu*
[1]University of Massachusetts Amherst, [2]Metropia

5:30 Machine Learning Methods to Predict Bus Travel Speeds and Analysis of the Impact of Different Predictive Variables
*Jan Berczely, Ricardo Giesen**
Pontificia Universidad Catolica de Chile

Choice between Metro and Taxi under Travel Time Variability

Gege JIANG[1], *Hong K. LO* [2,*]
Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology, Hong Kong, China
[*]Corresponding author: cehklo@ust.hk

1. Introduction

The modal choice problem has been investigated for many years, typically under deterministic travel conditions; not many studies have considered the effect of travel time variability endogenously. The context of this study is about the commute choices of travellers from home to work between two travel modes: metro services with a fairly constant or predictable travel time versus taxi that faces congestion and travel time variability. When choosing between these two modes, travellers may consider comfort, travel time, travel time variability and monetary cost. Among them, travel time reliability/variability has been increasingly acknowledged as an important factor affecting travel choices, as evidenced that unreliable travel times can disturb peoples' activities by causing unpredictable early or late arrival penalties (Batley and Ibáñez, 2012). Bhat and Sardesai (2006) made the early attempt to incorporate travel time reliability to modal choice in empirical experiments. They suggested that travel time as well as travel time reliability are both important indicators for level-of-service, which are highly valued by commuters, especially those with an inflexible work schedule. Li et al. (2010) addressed that people are willing to pay for improved travel time reliability. Chang (2010) proposed methods to evaluate travel time reliability in transport appraisals, which can be applied to multimodal research.

However, little attention has been paid on studying the theoretical underpinning of travel time reliability in the mode choice problem, let alone the pricing strategy for taxi. Many questions remain to be answered. How do travellers trade between travel time reliability and pricing, which in turn will affect their mode choices? Under the context of heterogeneous travellers, who will choose the metro service and who will choose the taxi? This study aims to answer these questions from an analytical perspective.

The structure of this abstract is as follows. The model assumptions and problem definitions are given in Section 2, first for the case of homogeneous travellers, then extended to the case of heterogeneous travellers. Due to page limitation, the results will be presented in the full paper.

2. Problem formulation

The problem is defined for N travellers commuting between home and work every day during morning peak hour. They choose between taking the taxi or the subway. The advantage of the subway is its predictable travel time, with little variability and lower fare, albeit it usually takes a longer travel time. Besides, it is less comfortable due to crowdedness. On the other hand, commuters taking cars have to endure congested travel time and travel time variability due to degradable road capacity and/or perhaps random demand. In this study, several assumptions are made to capture the travel behaviour and system performance:

(A1) The designed capacity for the road is $\bar{c}$, and the lower bound of the realized capacity is $\theta\bar{c}$. The probability distribution of the stochastic capacity $\bar{c}$ is known to all travellers, with the probability density function (*PDF*) $\phi(c)$. $\bar{c}$ follows a uniform distribution within interval $[\theta\bar{c},\bar{c}]$. Therefore, $\phi(c)=\dfrac{1}{(1-\theta)\bar{c}}$.

(A2) Travelers consider comfort, travel time, travel time variability and monetary cost in their choice mechanism.

(A3) The travel time variability for the subway is zero.

(A4) The waiting time and comfort loss of the subway is linear to the demand.

(A5) The comfort loss of taking the taxi is zero.

(A6) The operation cost of the car is linear to the actual travel time.

Previous researchers (e.g. Siu and Lo, 2008; Lo et al. 2006) has already derived the characteristics of travel time on a road with degradable capacity. Following their result, we adopt the bureau of public roads (BPR) link performance function:

$$t\left(N_c,c\right)=t_0\left[1+\gamma\left(\frac{N_c}{c}\right)^n\right] \tag{1}$$

where t_0 is the constant free flow travel time on the road. γ and n are deterministic parameters. c is the realized capacity on a specific day. The mean and variance of travel time can be derived as

$$E\left(t\right)=E\left(t_0\right)+\gamma t_0 E\left[\left(\frac{N_c}{c}\right)^n\right]=t_0+\gamma t_0 N_c^n E\left[\frac{1}{c^n}\right] \tag{2}$$

$$Var\left(t\right)=\gamma^2 t_0^2 N_c^n Var\left[\frac{1}{c^n}\right] \tag{3}$$

According to the assumption (A1), the mean and variance of $\frac{1}{c^n}$ are

$$E\left[\frac{1}{c^n}\right]=\int_{\theta\bar{c}}^{\bar{c}} c^{-n}\frac{1}{\left(1-\theta\right)\bar{c}}dc=\frac{1-\theta^{1-n}}{\bar{c}^n\left(1-\theta\right)\left(1-n\right)} \tag{4}$$

$$E\left[\frac{1}{c^{2n}}\right]=\int_{\theta\bar{c}}^{\bar{c}} c^{-2n}\frac{1}{\left(1-\theta\right)\bar{c}}dc=\frac{1-\theta^{1-2n}}{\bar{c}^{2n}\left(1-\theta\right)\left(1-2n\right)} \tag{5}$$

$$Var\left[\frac{1}{c^n}\right]=E\left[\frac{1}{c^{2n}}\right]-E^2\left[\frac{1}{c^n}\right]=\frac{1-\theta^{1-2n}}{\bar{c}^{2n}\left(1-\theta\right)\left(1-2n\right)}-\left[\frac{1-\theta^{1-n}}{\bar{c}^n\left(1-\theta\right)\left(1-n\right)}\right]^2 \tag{6}$$

2.1 Homogenous travellers

We start with homogenous travellers who have the same taste parameters. Based on the above assumptions, the disutility of taking the subway can be written as:

$$-U_s=\alpha_1 N_s+\alpha_2\mu_s+\alpha_5\tau_s \tag{7}$$

N_s is the number of travellers taking the subway. The first term $\alpha_1 N_s$ represents the sum of waiting time cost and comfort loss, which are both linear to the demand N_s . α_2 is the taste parameter for the mean travel time, which can be called the value of time (VOR). α_5 is the taste parameter for the monetary cost, which can be called the value of money (VOM). μ_s is the travel time for taking the subway. According to (A3), there is no variability cost for taking the subway. τ_s is the fixed monetary cost of taking the subway.

The disutility of taking the taxi service is:

$$-U_c=\alpha_2\mu_c+\alpha_3\sigma_c+\alpha_5\tau_c\left(t_c\right) \tag{8}$$

where

$$\mu_c=t_0+\gamma t_0 N_c^n\frac{1-\theta^{1-n}}{\bar{c}^n\left(1-\theta\right)\left(1-n\right)} \tag{9}$$

$$\sigma_c=\gamma t_0 N_c^n\sqrt{\frac{1-\theta^{1-2n}}{\bar{c}^{2n}\left(1-\theta\right)\left(1-2n\right)}-\left[\frac{1-\theta^{1-n}}{\bar{c}^n\left(1-\theta\right)\left(1-n\right)}\right]^2} \tag{10}$$

μ_c and σ_c are the mean and standard deviation of travel time t_c for taking the car. α_3 is the sensitivity parameter of σ_c, interpreted as value of reliability (VOR). $\tau_c\left(t_c\right)$ is the fee charged by the taxi service. The company can decide whether they charge a fixed fare every day or a variable fare according to the actual travel time on that day. We will discuss the two pricing scheme separately. The comfort loss of taking a car is zero.

For the taxi company, it aims to maximize the revenue. Scheme 1 is to charge a fixed fare every day. In equilibrium, all the travellers should have the same cost regardless of the mode they choose. Thus, the problem can be formulated as:

$$\max_{\tau_c} E\left[N_c\left(\tau_c-\beta t\right)\right]=\max_{\tau_c} N_c\left(\tau_c-\beta\mu_c\right) \tag{11}$$

$$\text{s.t} \quad -U_s=\alpha_1 N_s+\alpha_2\mu_s+\alpha_5\tau_s \tag{12}$$

$$-U_c=\alpha_2\mu_c+\alpha_3\sigma_c+\alpha_5\tau_c \tag{13}$$

$$N_s+N_c=N \tag{14}$$

$$\left(-U_s+U_c\right)N_c\ge 0 \tag{15}$$

$$\left(-U_s+U_c\right)N_s\le 0 \tag{16}$$

$$N_s,N_c\ge 0 \tag{17}$$

In (11), β is the operation cost per unit time for the taxi service. Equations (9) and (10) demonstrate that μ_c and σ_c are functions of N_c. So the car company determines an optimal point between the market share and the profit they can gain from each commuter.

For scheme 2, the company charges the fare by actual travel time. We assume that the fare and the actual travel time follows an implicit function of $\tau_c=f\left(t\right)$, $f^{'}\left(t\right)>0$. Substituting it into (11), we get the new objective function as:

$$\max_{f(t)} E\left[N_c\left(f\left(t\right)-\beta t\right)\right]=\max_{f(t)} N_c\left[E\left[f\left(t\right)\right]-\beta\mu_c\right] \tag{18}$$

In equilibrium, the expected cost of taking a car should be the same as taking the subway. Equation (13) now becomes

$$-U_c=\alpha_2\mu_c+\alpha_3\sigma_c+\alpha_5 E\left[f\left(t\right)\right] \tag{19}$$

The aim of the company is to find a charging policy to maximize its revenue. Now assume the optimal feasible fare of the first pricing scheme is τ_c^*. For each feasible solution of Scheme 2, $E\left[f\left(t\right)\right]$ is a constant, we denote it as τ'. Substituting $E\left[f\left(t\right)\right]=\tau'$ into (11)-(12), it is easily seen that Scheme 1 and Scheme 2 have the same expression. Therefore, the optimal feasible solution of Scheme 1 is also optimal feasible for Scheme 2 if and only if we manage the charging policy to have $E\left[f\left(t\right)\right]=\tau_c^*$. This result shows that under the context of linear disutility function of monetary cost, there is no difference between the two pricing schemes.

2.2 Travellers with same VOT and different VOR

Now suppose there are two kinds of travellers with same VOT but different VOR. We use α_{31} and α_{32} to denote the VOR of the commuters. For the first type, they take the variability into consideration (i.e. $\alpha_{31} > 0$), which is referred to as the risk-averse travellers. For the others, they don't care about the travel time variability (i.e. $\alpha_{32} = 0$), so they are risk-neutral ones. The number of risk averse and risk neutral travellers are N_1 and N_2 respectively. Denote the number of the first and second type of travellers taking the car as f_1 and f_2 respectively. $\hat{C}_c$ is the maximum cost for travellers choosing cars. Thus, the problem can be formulated as:

$$\max_{\tau_c} N_c\left(\tau_c - \beta\mu_c\right) \tag{20}$$

$$\text{s.t} \quad -U_p = \alpha_1 N_p + \alpha_2\mu_p + \alpha_5\tau_p \tag{21}$$

$$-U_{c1} = \alpha_2\mu_c + \alpha_3\sigma_c + \alpha_5\tau_c \tag{22}$$

$$-U_{c2} = \alpha_2\mu_c + \alpha_5\tau_c \tag{23}$$

$$f_1 + f_2 = N_c \tag{24}$$

$$N_p + N_c = N \tag{25}$$

$$\left(C_p - \hat{C}_c\right)N_p \le 0 \tag{26}$$

$$\left(C_{c1} - C_p\right) f_1 \le 0 \tag{27}$$

$$\left(C_{c2} - C_p\right) f_2 \le 0 \tag{28}$$

$$N_p,\, f_1,\, f_2 \ge 0 \tag{29}$$

It is easily seen that $C_{c1} > C_{c2}$ for $N_c > 0$. Therefore, risk averse travellers will first consider to take the subway. The result of this scenario will be beneficial for us to analyse the risk attitudes reflected by travellers' choices. We will extend this scenario to consider heterogeneous travellers. When the discretization interval is small enough, it can approximate the case with continually distributed VOR. The results will be instrumental in understanding the interplay between different pricing strategies and travellers with different VOT and VOR.

References

Batley, R. and Ibáñez, J. N. (2012). Randomness in preference orderings, outcomes and attribute tastes: An application to journey time risk. Journal of choice modelling, 5(3), 157-175.

Bhat, C. R. and Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B: Methodological, 40(9), 709-730.

Chang, J. S. (2010). Assessing travel time reliability in transport appraisal. Journal of Transport Geography, 18(3), 419-425.

Lo, H. K., Luo, X. W. and Siu, B. W. (2006). Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transportation Research Part B: Methodological, 40(9), 792-806.

Siu, B. W., & Lo, H. K. (2008). Doubly uncertain transportation network: degradable capacity and stochastic demand. European Journal of Operational Research, 191(1), 166-181.

Multi-Cycle Optimal Taxi Routing with E-Hailing

Xinlian Yu* Song Gao† Hyoshin Park‡ Xianbiao Hu§

1 Introduction

1.1 Motivation

Taxis play an important role in providing on-demand mobility in the urban transportation system. Compared to other forms of public transportation, the advantages of taxis include speediness, privacy, comfort, door-to-door service and 24/7 operations. Traditionally, vacant taxis cruise on roads searching for customers. In recent years, thanks to the advance of smartphone technology, e-hailing applications (e.g., Uber, Lyft, and Didi) are widely adopted by the drivers to receive requests from nearby customers. An occupied taxi usually takes a direct route to the customer's destination, not unlike regular commuters. However, there is no guarantee that the driver can find a new customer after dropping off the previous customer at the destination. Vacant taxis cruising on roads not only result in wasted gas and time for taxi drivers but also generate additional traffic in a city. Therefore, how to improve the utilization of taxis is of importance to both taxi drivers and the society.

In an earlier study by the co-authors (Hu et al., 2012), a dynamic programming model of routing vacant taxis was proposed to depict the decisions at intersections according to the passenger arrival rate. However, the expected search time is only minimized for the next customer, which might be inefficient in the long run. For example, driving to the airport might not minimize the search time for the next customer, but it brings in a higher chance of a long trip for the next customer and thus the profit is higher in a long period. For this reason, experienced taxi drivers would not simply make their customer-search decisions depending on the current searching time/profit, but would also consider the subsequent possible states that could be encountered. In this study, an optimal taxi routing problem is investigated for a single taxi that accounts for multiple cycles of pick-up and drop-off into the future.

1.2 Literature Overview

Since the early 1970's, a large number of studies on taxis have been conducted. See Salanova et al. (2011) for a recent review. Static or myopic taxi routing modeling and optimization is beyond the

*University of Massachusetts, Amherst. xinlianyu@umass.edu
†University of Massachusetts, Amherst. sgao@umass.edu
‡University of Massachusetts Amherst. hyoshinpark@umass.edu
§Metropia. xb.hu@metropia.com

scope of this study.

For modeling multi-period taxi services, Yang et al. (2005) presented a spatially aggregated taxi service equilibrium model with endogenous service intensity. However its main purpose is to understand the demand-supply interaction instead of improve taxi utilization. Wong et al. (2015) developed a sequential logit-based vacant taxi behavior model predicting searching paths as a sequence of choices of adjacent zones. The destination choice is restricted to adjacent zones, while an experienced driver chooses destinations over the whole network (e.g, going back to airports and major business districts after dropping off customers in a remote residential area). Furthermore, the search behavior model is based on zones and cells instead of the road network.

Table 1: Summary of selected literature

Problem	**Studies**	**Model features**
Equilibrium of taxi supply and demand	Yang et al. (2005)	period-specific taxi customer demand as a function of waiting time and taxi fare
Vacant taxi searching behavior	Wong et al. (2015)	logit-based customer searching paths; restricted to one searching trip
Taxi demand prediction	Moreira-Matias et al. (2012)	time series forecasting model
	Qian and Ukkusuri (2015)	geographically weighted regression (GWR)
	Yuan et al. (2011) Hwang et al. (2015)	mining historical GPS data through machine learning
Non-myopic VRP	Thomas and White (2004)	One-stage look-ahead
	Mitrović-Minić et al. (2004)	double-horizon heuristic
	Ferrucci et al. (2013)	rolling horizon heuristic;future requests arrive in a time-space Poisson process

Models and algorithms developed for non-myopic vehicle routing problem (VRP) under uncertainty with look-ahead policies and rolling horizons (e.g., Mitrović-Minić et al., 2004; Thomas and White, 2004; Ferrucci et al., 2013) might provide insights for taxi routing problems in terms of accounting for future unknown demand and efficient solution algorithms. It is however recognized that the taxi problem is different. In a typical VRP, the service of a customer does not bring the vehicle to another location, while a taxi does and the destination is not known until the request is taken. This significantly increases the geographic spread of taxi movements. In addition, a taxi (without carpooling service) can serve only one quest at one time and a new request does not come up until the old request is finished (unless a dispatcher is sending request during the previous ride).

Accounting for future states in taxi searching behavior requires sound models of geographic and temporal distributions of taxi demand. Several methods have been proposed to predict taxi demand distribution (see ”Taxi demand prediction” in Table 1), which could be combined with the optimal taxi routing model.

1.3 Contributions

Define a trip cycle consisting of the vacant taxi trip from the destination of the previous occupied trip to the pick-up place for the next customer, and the subsequent occupied trip from pick-up place for transporting the customer to his/her destination. A non-myopic, multi-cycle taxi routing optimization problem is studied with the following intended contributions:

- The optimization problem takes into account future states by considering both the intensity of future customer demand and their destinations over multiple cycles.
- Instead of zone/cell-based movements, the routing decisions are based on the physical road network, which enables potentially better understanding of taxi drivers' behaviors as well as more practical recommendations for taxi drivers.
- Practical implementations are proposed to solve the multi-cycle optimal taxi routing problem in a reasonable time. The solution will be compared with observed searching behaviors from GPS trajectories in a mega city to demonstrate the advantage of the multi-cycle approach.

2 Formulation of the Multi-Cycle Optimal Taxi Routing Problem

2.1 Taxi Movements in a Network

A taxi travels in a traffic network $G = (N, A)$. N is the set of nodes and A the set of links. There is at most one directional link, a, from the source node i to sink node j. $A(i)$ is the set of downstream links of i, and $B(j)$ the set of upstream links of j. The taxi is actively searching for, or carrying passengers during a time horizon with discretized time intervals $\{0, 1, \ldots, M-1\}$. For simplicity, link travel times are assumed time-independent and deterministic, but the formulation can be generalized to account for time-dependent and/or stochastic link travel times. The length of a time interval δ is equal to the shortest link travel time, and travel time for link a is represented as a non-negative integer, $\tau_a = \lfloor x_a/\delta \rfloor$, where x_a is the original travel time for link a.

The routing problem is meaningful only when a taxi is for hire. The state of a taxi, s, is described by node i and time interval t. The action set for state s is the set of outgoing links $A(i)$. For a given state s and action $j \in A(i)$, two types of transition to a new state s' could happen. 1) The taxi is not matched with any passenger when traversing link $a = (i, j)$, and s' is associated with node j. 2) The taxi is matched with a passenger when traversing link a, and s' is associated with the destination node of the passenger, k. By definition, any state associated with the last time period $M-1$ is a terminal state. To calculate state transition probabilities, the passenger matching probability on a link (Section 2.2) and passenger destination probabilities (Section 2.3) are needed.

2.2 Passenger Matching Probability on a Link

Passengers arrive at link a following a one-dimensional space-time Poisson process with rate λ_a per hour per mile. For modeling convenience, these are simplified as time Poisson processes at nodes, and the arrival rate at node j (per hour), $\lambda_j = \sum_{a \in B(j)} \lambda_a l_a$, where l_a is the length of link a. Arrivals at different nodes are independent. In practice, demand rate λ_j is often approximated by observed met demand rate. Statistical analysis can be carried out to build a predictive model for the demand rate as a function of built environment variables (e.g., residential density, and employment by business type such as hotel and nightclub), time of day, and weather condition (Phithakkitnukoon et al., 2010; Moreira-Matias et al., 2012).

When e-hailing is used, the nearest vacant taxi to a passenger gets matched to the passenger. Vacant taxis around node n at any given point of time follow a two-dimensional spatial Poisson distribution with density γ_n. For simplicity, hypothetical right-angle travel is assumed. For a given node n, the probability of a vacant taxi r miles away (based on right-angle travel) being the nearest vacant taxi is the probability of no vacant taxi in a square rotated at 45 degree centered at node n with area equal to $2r^2$, namely,

$$P_n(r) = \exp(-2\gamma_n r^2). \tag{1}$$

Consider a vacant taxi with e-hailing traversing link a. It gets matched with a passenger at node h if a passenger arrives at node h during the traversal time τ_a and the vacant taxi is the nearest. The matching probability, $p_{a,h}$, is the product of the probability that a passenger arrives at node h within τ_a and the probability that the vacant taxi is the nearest to node h, namely,

$$p_{a,h} = (1 - \exp(-\lambda_h \tau_a)) \exp(-2\gamma_h \mathcal{L}^2_{h\to a}), \tag{2}$$

where $\mathcal{L}_{h\to j}$ is the right-angle distance from node h to link a, which can be approximated as the distance to the middle point of link a. Note the pickup nodes can be restricted to a subset of the nodes to model picking up along the roads without e-hailing.

2.3 Passenger Destination Probabilities

The probability of a passenger picked up at node h having node k as the destination, $p_{h\to k}$, can be approximated by the observed fraction of passengers picked up at node h going to k. When no passenger pick-up is observed at node h, the probability is undefined. To resolve this problem, the study area is divided into zones such that any zone has strictly positive number of pick-ups. Let node h be in zone $\mathcal{H}$ and node k in zone $\mathcal{K}$. Assume each node in zone $\mathcal{K}$ has equal probability of being the destination node, and the destination probability is

$$p_{h\to k} = \begin{cases} \dfrac{p_{\mathcal{H}\to\mathcal{K}}}{m_{\mathcal{K}}}, & \forall \mathcal{H} \neq \mathcal{K} \\ \dfrac{p_{\mathcal{H}\to\mathcal{K}}}{m_{\mathcal{K}} - 1}, & \forall \mathcal{H} = \mathcal{K}, \forall h \neq k \\ 0, & \forall h = k \end{cases}, \tag{3}$$

where $p_{\mathcal{H}\to\mathcal{K}}$ is the probability of a passenger picked up in zone $\mathcal{H}$ having zone $\mathcal{K}$ as the destination zone, and $m_{\mathcal{K}}$ is the number of nodes in zone $\mathcal{K}$. The equal probability assumption can be easily relaxed.

2.4 State Transition Probabilities and the Optimization Problem

For a given state $s = (i, t)$ and action $a \in A(i)$ with a sink node j, the transition probability, $p_{ss'|a}$ with $s' = (i', t')$, is defined as follows:

$$p_{ss'|a} = \begin{cases} 1 - \sum_{h\in N} p_{a,h}, & \text{if } i' = j, t' = t + \tau_a \\ \sum_{h|t'=t+\tau_a+\mathcal{T}_{j\to h}+\mathcal{T}_{h\to i'}} p_{a,h}\, p_{h\to i'}, & \text{if } (i' = j, t' \neq t + \tau_a) \text{ or } (i' \neq j) \end{cases}, \tag{4}$$

where $\mathcal{T}_{j\to h}$ ($\mathcal{T}_{h\to i'}$) is the shortest path travel time from node j to h (h to i'). In the first case, no passenger is matched along link a, and the taxi arrives at the sink node j at time $t+\tau_a$. In the second case, a passenger from node h with destination i' is matched, and the taxi arrives at node i' after picking up the passenger from node h and carrying the passenger from h to i', both following shortest paths. The probabilities are summed over all possible h such that the arrival time at the destination i' is t'.

It follows that the immediate payoff of going from state s to s' given action a can be written as follows:

$$g_{ss'|a} = \begin{cases} -\alpha\tau_a, & \text{if } i'=j, t'=t+\tau_a \\ \dfrac{\sum_{h|t'=t+\tau_a+\mathcal{T}_{j\to h}+\mathcal{T}_{h\to i'}}(-\alpha(t'-t)+\beta\mathcal{T}_{h\to i'})p_{a,h}p_{h\to i'}}{\sum_{h|t'=t+\tau_a+\mathcal{T}_{j\to h}+\mathcal{T}_{h\to i'}} p_{a,h}\, p_{h\to i'}}, & \text{if } (i'=j, t'\neq t+\tau_a) \text{ or } (i'\neq j) \end{cases}, \tag{5}$$

where α is the taxi operating cost per hour, and β is the taxi revenue per hour.

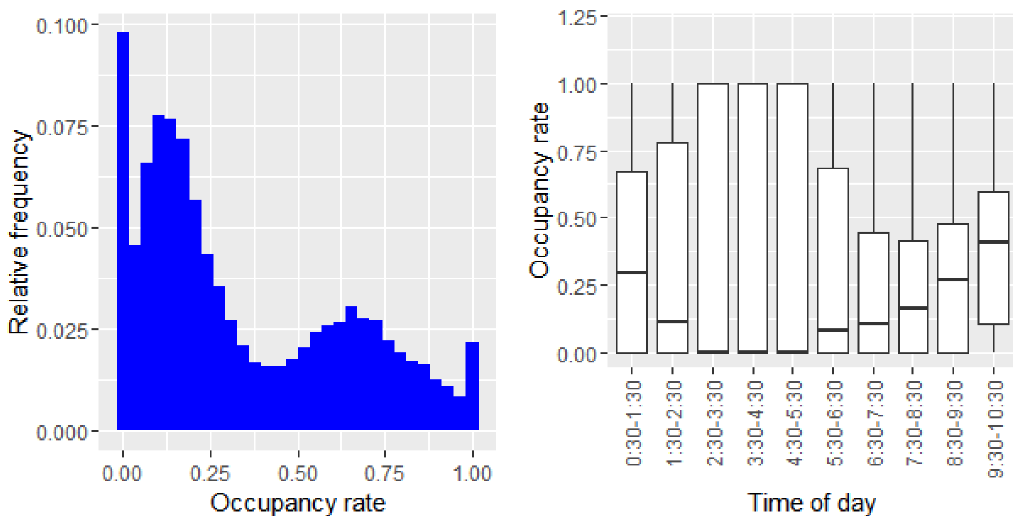
Let $V(s)$ denote the optimal expected payoff starting from state s. The taxi driver chooses the action at each state s to maximize the expected payoff that is the sum of the expected immediate payoff and the expected downstream payoff, which is the expectation of the payoff over all possible next state s'. The optimal solution is obtained by solving the Bellman equation (Bellman, 1957) as follows:

$$V(s) = \begin{cases} \max\limits_{a\in A(i)} \sum\limits_{s'}(g_{ss'|a}+V(s'))p_{ss'|a}, & \forall s|t<M-1 \\ 0, & \forall s|t=M-1. \end{cases} . \tag{6}$$

3 Case Study and Anticipated Results

We will evaluate our model using a large number of historical GPS trajectories generated by urban taxis in Shanghai, China, in the month spanning September 1-30, 2014. Figure 1 reveals some statistics of 13,544 taxicabs from 0:00 to 10:30am on a typical Monday, September 8, 2014. As shown in Figure 1(a), the distribution of taxi occupancy rate (the quotient between the occupied time and the whole working period) exhibits a bi-modal pattern with two peaks: 0-0.25 and 0.5-0.75. The occupancy rate of taxi drivers varies widely and most taxi drivers have a pretty low occupancy rate between 0 to 0.25. Figure 1(b) further shows the boxplots of occupancy rate over time. During peak hours, the gap between the high-occupancy rate drivers and low-occupancy rate drivers is less obvious than that during off-peak hours. More drivers can find passengers easier during peak hours. However, it is also difficult to obtain a high occupancy rate due to the increasing competition between drivers during peak hours. The occupancy rate during 2:30-7:30am is in general low, yet more drivers could reach an occupancy rate of 1 during 3:30-5:30am due to less competition, longer taxi trips and more experienced drivers who know the places where he/she can pick up passengers quickly during this time.

The main challenge in solving the multi-cycle optimal routing problem in real-world networks lies in calculating the cost-to-go with a large number of possible future states. Approximate Dynamic Programming (ADP) provides a potential tool for calculating the future impact of a current



(a) Distribution of occupancy rate

(b) Occupancy rate over time

Figure 1: Statistics on the occupancy rate

decision. ADP approximates the value function based on some methods (e.g. reinforcement learning, Q-learning and simulation), and ultimately avoids the evaluation of all possible states. To further cope with the scalability problem, a rolling horizon approach where the planning time horizon is limited in length (e.g, a few hours long) will be also be investigated. We will show that the multi-cycle optimization problem can be solved in a reasonable time, while providing significantly better solutions in the long run by comparing the model results against observed searching behaviors.

References

Bellman, R. (1957). A markovian decision process, *Technical report*, DTIC Document.

Ferrucci, F., Bock, S. and Gendreau, M. (2013). A pro-active real-time control approach for dynamic vehicle routing problems dealing with the delivery of urgent goods, *European Journal of Operational Research* **225**(1): 130–141.

Hu, X., Gao, S., Chiu, Y.-C. and Lin, D.-Y. (2012). Modeling routing behavior for vacant taxicabs in urban traffic networks, *Transportation Research Record: Journal of the Transportation Research Board* **2284**: 81–88.

Hwang, R.-H., Hsueh, Y.-L. and Chen, Y.-T. (2015). An effective taxi recommender system based on a spatio-temporal factor analysis model, *Information Sciences* **314**: 28–40.

Mitrović-Minić, S., Krishnamurti, R. and Laporte, G. (2004). Double-horizon based heuristics for the dynamic pickup and delivery problem with time windows, *Transportation Research Part B: Methodological* **38**(8): 669–685.

Moreira-Matias, L., Gama, J., Ferreira, M. and Damas, L. (2012). A predictive model for the passenger demand on a taxi network, *2012 15th International IEEE Conference on Intelligent Transportation Systems*, IEEE, pp. 1014–1019.

Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A. and Ratti, C. (2010). Taxi-aware map: Identifying and predicting vacant taxis in the city, *International Joint Conference on Ambient Intelligence*, Springer, pp. 86–95.

Qian, X. and Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using gps data, *Applied Geography* **59**: 31–42.

Salanova, J. M., Estrada, M., Aifadopoulou, G. and Mitsakis, E. (2011). A review of the modeling of taxi services, *Procedia-Social and Behavioral Sciences* **20**: 150–161.

Thomas, B. W. and White, C. C. (2004). Anticipatory route selection, *Transportation Science* **38**(4): 473–487.

Wong, R., Szeto, W. and Wong, S. (2015). A two-stage approach to modeling vacant taxi movements, *Transportation Research Part C: Emerging Technologies* **59**: 147–163.

Yang, H., Ye, M., Tang, W. H.-C. and Wong, S. C. (2005). A multiperiod dynamic model of taxi services with endogenous service intensity, *Operations research* **53**(3): 501–515.

Yuan, J., Zheng, Y., Zhang, L., Xie, X. and Sun, G. (2011). Where to find my next passenger, *Proceedings of the 13th international conference on Ubiquitous computing*, ACM, pp. 109–118.

Machine Learning Methods to Predict Bus Travel Speeds and Analysis of the Impact of Different Predictive Variables

Jan Berczely; Ricardo Giesen*
jberczel@uc.cl, Giesen@ing.puc.cl
Department of Transport Engineering and Logistics
Pontificia Universidad Católica de Chile
(*) Corresponding author

A critical indicator for most public transportation agencies is to increase the number of passengers in the system. To achieve this goal, the level of service offered to the users must be maintained or improved. This is very hard, since traffic congestion is normally increasing which affects bus travel times. As a result, bus travel times are higher and less reliable, which makes harder to predict travel times and avoid bunching. Being able to accurately predict bus travel speeds and update this prediction with real-time information could improve the quality and reliability of the information given to users, since it helps them to plan their routes, and to reduce the anxiety of waiting without knowing the time of arrival of their bus. In addition, for operators, better speed predictions increase the effectiveness of control schemes to avoid bunching and allow them to optimize the number of buses required to operate a service.

Machine learning models have seized high popularity because of their good results and ability to easily adapt to the requirements of the modeler. As a consequence, several studies have been reported in predicting travel speed/time of buses. There are many models of statistical learning to predict bus travel speeds, and there is no consensus regarding the best method. In this paper, we work with three models that have been widely used in this field: (i) Multiple Linear Regression (MLR), (ii) Support Vector Machines (SVM) and (iii) Neural Networks (ANNs). SVM and ANN were chosen because they are methods that have reported very good results, see for example Jeong & Rilett (2004), Yu, Lam & Tam (2011), and Zheng, Zhang & Feng (2012). However, SVM and ANN are black box models, in which is difficult to understand the impact of each input variable in the predictions. This is why MLR was incorporated, since although it has not reported as good results as the other two (Jeong & Rilett, 2004; Yu et al., 2011), MLR models are easy to calibrate and interpret. Thus using MLR, we can clearly see the contribution of each of the input variables in the predictions.

This paper has two objectives. The first one is to compare the performance of three proposed models Multiple Linear Regression, Support Vector Machines and Neural Networks, in comparison to two benchmark models that use averages of historical and real-time information. To do this, real data from three services of buses of the city of Santiago, Chile are used. The second objective seeks to determine which explanatory variables are more or less significant in the outcome of the model. Julio et al (2016) propose a series of models to predict bus speeds in the city of

Santiago, however, they only use as input variables data from the bus position reported every 30 seconds by GPS of each bus to make the predictions. Our analysis includes variables such as bus speed reported by GPS, passengers boarding and alighting at different bus stops, bus load, infrastructure and environmental factors such as weather conditions.

Figure 1 and Table 1 present a description of the services used to study the performance of different predictive models. Service 212 is a classical trunk service with most of its route in segregated corridors and exclusive lanes, service 203 is also a trunk service with have only half of its route in segregated corridors and exclusive lanes, and service C04 is a feeder service with a shorter route and smaller buses running in mixed traffic without priority.

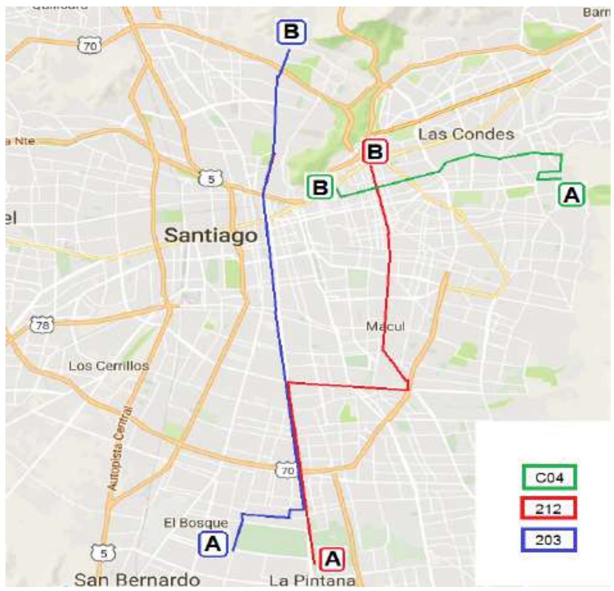


Figure 1 - 1. Map of Services Studied 212, 203 and C04

Service	Direction	Total Length (Km)	Average Speed (km/hr)	Average Headway (min)	Average Distance between Stops (m)
212	Outbound	28	22.3	7.3	350
203	Outbound	30	21.9	5.6	315
C04	Inbound	12	20.8	17.7	270

Table 1. General Characteristics of the Services 212, 203 and C04.

Tables 2 to 4 show the results obtained by the different models when predicting bus speeds in the three services studied. In terms of RMSE ANN is the best model. Even though MLR gets slightly worse results, on average 0.2% worse, MLR models are easy and fast to calibrate, so we believe that they should not be ruled out as alternative if you consider that in a system such as Santiago there are more than 350 lines, and twice as many services (DPTM, 2013).

	HB	RTB	MLR	ANN	SVM
RMSE (km/hr)	6.95	6.12	5.47	5.46	5.57
MAPE (%)	20.42	15.39	17.14	17.19	16.35

Table 2. Performance measures for 212-outbound service

	HB	RTB	MLR	ANN	SVM
RMSE (km/hr)	7.05	6.37	5.78	5.76	5.85
MAPE (%)	21.28	16.91	18.79	18.73	17.75

Table 3. Performance measures for 203-outbound service

	HB	RTB	MLR	ANN	SVM
RMSE (km/hr)	5.62	4.98	4.42	4.40	4.51
MAPE (%)	18.81	13.63	14.65	14.35	14.40

Table 4. Performance measures for C04-inbound service

In the three bus services analyzed, the model that showed the best results, in terms of the RMSE is Neural Networks, followed by Multiple Linear Regression and then by the Support Vector Machines. In all cases, machine learning models exceeded benchmark models, with a performance that varies between 10% and 25% decrease in the error. These results highlight the value of using real-time information and machine learning methods to improve the accuracy of predictions.

Regarding the explanatory variables involved, we concluded that only speed variables have a relevant impact, whilst the addition of the other analyzed variables only had a minor effect (less than 2% reduction of errors). It seems that speed variables have implicit, in their value, the effect of the other variables, such as demand, infrastructure and environmental factors.

REFERENCES

(DPTM), D. d. (2013). *Informe de Gestión 2013.*

Coffey, C., Pozdnoukhov, A., & Calabrese, F. (Noviembre de 2011). Time of arrival predictability horizons for public bus routes. *In Proceedings of the 4th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, 1-5.

Cortés, C. E., Gibson, J., Gschwender, A., Munizaga, M., & Zúñiga, M. (2011). Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies, 19(4)*, 695-707.

Du, L., Peeta, S., & Kim, Y. H. (2012). An Adaptive Information Fusion Model to Predict the Short-Term Link Travel Time Distribution in Dynamic Traffic Networks. *Transportation Research part B 46*, 235-253.

Gurmu, Z. K., & Fan, W. D. (2014). Artificial neural network travel time prediction model for buses using only GPS data. *Journal of Public Transportation, 17(2)*.

Jeong, R., & Rilett, L. (2004). Bus arrival time prediction using artificial neural network model. *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference* (págs. 988-993). IEEE.

Julio, N., Giesen, R., & Lizana, P. (2016). Real-time prediction of bus travel speeds using traffic shockwaves and machine learning algorithms. *Research in transportation economics*.

Kumar, V., Kumar, B. A., Vanajakshi, L., & Subramanian, S. C. (2014). Comparison of model based and machine learning approaches for bus arrival time prediction. *Proceedings of the 93rd Annual Meeting. Transportation Research Board*, 14-2518.

Mazloumi, E., Rose, G., Currie, G., & Moridpour, S. (2011). Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. *Engineering Applications of Artificial Intelligence, 24(3)*, 534-542.

Rahman, M., Wirasinghe, S., & Kattan, L. (2016). The effect of time interval of bus location data on real-time bus arrival estimations. *Transportmetrica A: Transport Science, 12:8*, 700-720.

Wang, L., Zuo, Z., & Fu, J. (2014). Bus Arrival Time Prediction Using RBF Neural Networks Adjusted by Online Data. *Procedia-Social and Behavioral Sciences*(138), 67-75.

Xinghao, S., Jing, T., Guojun, C., & Qichong, S. (2013). Predicting bus real-time travel time basing on both GPS and RFID data. *Procedia-Social and Behavioral Sciences, 93*, 2287-2299.

Yu, B., Lam, W. H., & Tam, M. L. (2011). Bus arrival time prediction at bus stop with multiple routes. *Transportation Research Part C: Emerging Technologies*, *19(6)*, págs. 1157-1170.

Zheng, C. J., Zhang, Y. H., & Feng, X. J. (2012). Improved iterative prediction for multiple stop arrival time using a support vector machine. *Transport, 27(2)*, 158-164.

Zhong, S., Hu, J., Ke, S., Wuang, X., Zhao, J., & Yao, B. (2015). A Hybrid Model Based on Support Vector Machine for Bus Travel-Time Prediction. *Traffic & Transportation, 27*, 291-300.

Multimodal Transportation Services

FA1: Transit Fare and Revenue

Friday 8:30 – 10:30 AM

Session Chair: Hai Yang

8:30 Ticket Portfolio Planning for Urban and Regional Public Transport Networks
*Jörn Schönberger**
Technische Universität Dresden

9:00 Optimal Discount Policies for Transit Agencies: The Case of Pass Programs and Loyalty Programs
[1]Mehdi Nourinejad,[2] Amir Gandomi, [3]Joseph Y. J. Chow, [1]Matthew J. Rooda*
[1]University of Toronto, [2]Ryerson University, [3]New York University

9:30 Pricing for a Last Mile Transportation System
[1]Hai Wang, [2]Yiwei Chen*
[1]Singapore Management University, [2]Singapore University of Technology and Design

10:00 A Fare-Reward Scheme for Commuters in Transit Bottleneck
Yili Tang, Hai Yang*
The Hong Kong University of Science and Technology

Ticket Portfolio Planning for Urban and Regional Public Transport Networks

Jörn Schönberger

Technical University of Dresden, Chair of Transport Services and Logistics, Faculty of Transportation and Traffic Sciences, Germany,

joern.schoenberger@tu-dresden.de

A careful planning of public transport operations is the key aspect of a successful commuting and transfer system. In the last decades we have seen significant progress in network design, route planning, and timetabling and crew rostering as well as delay and disruption management. All these approaches have led to an increase of efficiency in the provided daily operations of commuter service providing companies.

The planning of operations contributes to keep costs on an acceptable level in order to cope with short and limited budgets. These budgets are mainly provided by two sources. First, transfer payments are received from service ordering authorities like municipal or regional administrations. Second, customer revenues from ticket sales form the second source of incoming money for a commuter operator.

Transfer payments are often coupled with the provided services and their actual reliability. Here, the operations planning is adequate to contribute to reliable services that protect the income of the service providing companies from the transfer payment. However, this part of income is typically fixed and cannot be increased due to appropriate planning.

With respect to the increase of the total budget the increase of the revenues from sold tickets seems to be a good idea. In some application contexts like the airline business, tourism transport or long-distance train services revenue management techniques like price discrimination, capacity control, overbooking or dynamic pricing have proven to be able to contribute to the generation of additional revenues. Here, the central mechanisms behind the aforementioned techniques are (i) the estimation of the maximal willingness to pay of an individual customer and (ii) the acquisition of the major part of the customer's surplus by setting an individual price for each sold ticket.

In the commuter business, the implementation of such approaches is impossible. It is therefore interesting to identify possible planning approaches that enables a service proving company to raise the amount of sold tickets and/or the gained revenues from the ticket fare. This presentation aims at providing initial insights into ideas to enable revenue planning in the commuter business.

A commuter ticket, independently whether it is a single trip ticket or a period ticket, is the „product" sold by a public transport company. As observed for the manufacturing of material products, the definition of such a product requires several preparatory steps. Although there is no „common theory" documented in the literature, one can group the decision tasks associated with the specification of offered commuter tickets into three categories („layers") which have to be executed consecutively.

The first step is the definition of fare zones. A set of fare zones partitions the complete service area into portions (the fare zones). For each origin-to-destination trip in the covered service area the involved (e.g. visited) zones are used to determine the money a traveler has to pay to the service provider for this trip. Several papers discuss mathematical models for fare zone definition tasks. These models are identified as quite complex so that sophisticated algorithms are needed to identify appropriate fare zones.

The second step is the ticket portfolio compilation. This decision problem addresses the task which tickets are offered to the customers. A ticket gives the traveler the right to travel to a certain extent and/or for a certain time in the complete network or in a specific part of the service network. Typically, the allowed part of the network is expressed by listing a subset of the zones determined in the fare zone definition step. In the literature, we found some papers that evaluate the offered ticket portfolio but to the best of the author's knowledge, there are no papers addressing the planning of the ticket portfolio composition.

The third and final product specification step is ticket pricing. Here, a price is assigned to each ticket contained in the ticket portfolio. The major challenges in commuter ticket pricing are (i) political and social needs must be considered (ii) price sensitivity of customers who will not buy a too expensive ticket due to alternative modes of transport (iii) limited transport capacities that do not allow a too low ticket price and (iv) to ensure the consistency between prices assigned to different tickets, e.g. a week ticket must not cost more than seven times the price for a one day ticket. We find some papers that provide and evaluate models to identify adequate fares for tickets in a portfolio.

After having structured the area of product specification in urban and regional public transport it turns out that there are several contributions to the planning of fare zone systems (the first layer). Furthermore, several contributions to the determination of fares for offered tickets are available (the third layer). However, more or less nothing is contributed to the portfolio planning for tickets in commuter systems. Therefore, there is a research gap associated with the second layer.

In order to start closing the detected research gap we propose mathematical optimization models that support the planning of optimal ticket portfolios to be offered by commuter agencies. In these models, the decisions to be made correspond to the selection of possible tickets for inclusion into a ticket portfolio. Within these models we test different planning goals with respect to their appropriateness and the need to establish necessary constraints. Furthermore, we investigate the influence of restrictions imposed in order to make the resulting ticket portfolio as easily understandable for customers.

After we have proven the suitability of the initial models we propose several real-world-borrowed extensions and modifications. In particular, we incorporate a time-based price discrimination scheme. It enables the introduction of peak time tickets as well as low demand period discounted tickets. The proposed models are evaluated within computational experiments.

The proposed decision models represent a first approach to equip commuter agencies with a decision support tool for the compilation of the „best fitting“ ticket portfolio. Nevertheless, a lot of issues as well as mechanism we are observing in commuter system ticket portfolios are still uncovered. Among these issues, we want to refer to multi-trip tickets and season tickets (daily, weekly, monthly or annually) as well as special purpose tickets for the elderly, handicapped or persons with low income. Transferability issues that allow the transfer of a ticket between different customers as well as the allowance of free rides for accompanying persons represent other important but still unsolved challenges in the commuter ticket portfolio compilation.

Optimal Discount Policies for Transit Agencies: The Case of Pass Programs and Loyalty Programs

Mehdi Nourinejad*[a], Amir Gandomi[b], Joseph Y.J. Chow[c], Matthew J. Roorda[d]

[a]*Department of Civil Engineering, University of Toronto, Toronto, Canada*
mehdi.nourinejad@mail.utoronto.ca
[b]*Ted Rogers School of Management, Ryerson University, Toronto, Canada*
agandomi@ryerson.ca
[c]*Department of Civil and Urban Engineering, New York University, New York, United States*
joseph.chow@nyu.edu
[d]*Department of Civil Engineering, University of Toronto, Toronto, Canada*
roordam@ecf.utoronto.ca

Abstract

The proliferation of smart cards in public transportation has paved the way for successful implementation of two prominent discount policies: pass programs and loyalty programs. While pass programs have been around as early as the 1970s, loyalty programs are only now gaining unprecedented popularity in public transportation. In a loyalty program, riders get a discount on their fare if they complete a given number of trips within a time period (e.g., a month). Our review of several mass transit agencies shows that transit agencies are not unanimous in their choice of discount policy. While some offer only the pass program, others offer the loyalty program, and a few offer both. In this paper, we derive the optimal pass and loyalty program, and we investigate if one is superior to the other in terms of social welfare and profit. We find that each program has a unique impact on the transit agency. The pass program is superior to the loyalty program for public transit agencies because (i) it offers a higher social welfare, and (ii) social welfare and profit are maximized at the same pass price, thus indicating that agencies do not sacrifice profit for social welfare. The loyalty program, on the other hand, is beneficial for private agencies such as Uber and Lyft because it generates a higher profit than the pass program when riders have more non-mandatory than mandatory trips. We develop a simulation model to account for several sources of uncertainty including user heterogeneity. The simulation model validates our earlier findings from the analytical expressions and provides several insights as well.

1. Introduction and motivation

Discount programs have gained unprecedented popularity in public transportation as they are known to improve rider convenience and operator profitability. As early as the 1970s, transit agencies have offered pass programs that grant riders unlimited trips for a given planning period (e.g. a month). Pass programs are known to decrease auto use, improve accessibility, and reduce passenger boarding time by up to 25% (Lachapelle and Frank, 2009; Cervero, 1990). In addition to pass program, the proliferation such fare payment technologies as smart cards has paved the way for more sophisticated discount policies including loyalty programs (Li et al., 2009; Pelletier et al., 2011; Chow, 2014). Loyalty programs offer riders a discount on their transit fare if they complete a predefined number of trips within a given period of time (e.g. a month). For example, PRESTO, a smart-card fare payment company in Canada, offers riders a 87.75% discount if they make more than 36 trips per month via Go-Transit (i.e., any trip beyond the 36^{th} trip is discounted) (Go Transit, 2016). It is not yet clear if pass programs are superior to loyalty programs or the other way around. A synopsis of current transit agencies in Ontario, Canada, shows that agencies are not unanimous in their choice of discount policy. As an example, Table 1 shows that Brampton Transit is adopting a pass program, whereas MiWay is adopting a loyalty program, and Burlington Transit is implementing both. The motivation of this study is to investigate which discount policy is ideal under various operating conditions. We find the optimal design of each policy and examine their impact on social welfare and profit.

2. The model

2.1. Problem setting

Consider a homogeneous group of riders where each rider makes m mandatory and n non-mandatory trips during a given planning horizon (e.g. a week or a month). Mandatory trips are trips made by travelers in the absence of any discount policy and non-mandatory trips are those that are potentially made when a discount is provided. In the absence of a discount program, each rider pays a *fare* of f [dollars] per ride and receives a marginal utility per trip. The mandatory trips, compared to the non-mandatory trips, yield a higher marginal

Table 1: Discount programs of Canadian public transportation agencies. Monetary values are in Canadian dollars. Entries were retrieved from agency websites.

Agency	Fare per trip	Pass Program	Loyalty Program
Hamilton Street Railway	\$3.00	\$ 101.20 / month	free after 11 trips per week
Burlington Transit	\$2.70	\$97.00 / month	free after 36 trips per month
Oakville Transit	\$2.80	\$115 / month	free after 35 trips per month
MiWay, Mississauga	\$2.90	-	free after 12 trips per week
Go Transit	\$5.30	-	87.75% discount after 36 trips per month
Brampton Transit	\$2.90	\$120 / month	-
OC Transpo, Ottawa	\$3.00	\$105.75 / month	-
Durham Region Transit	\$3.05	\$115 / month	planned
Toronto Transit Commission	\$2.90	\$141.5 / month	planned
York Region Transit	\$3.40	\$140.00 / month	planned

utility because they are valued more by riders. The marginal utility function is denoted by $u(t)$ and presented in Fig.1. In this figure, the trips are laid out on the x-axis in the order of their marginal utility so that the m mandatory trips appear before the n non-mandatory trips. We assume in Fig.1 that the marginal utility of the m^{th} (final) mandatory trip is f [dollars] (i.e., $u(m) = f$). This implies that when no discount is offered, each rider makes only m trips during the planning horizon. We also assume that $u(t)$ is linearly decreasing, although this assumption does not pose a serious restriction.

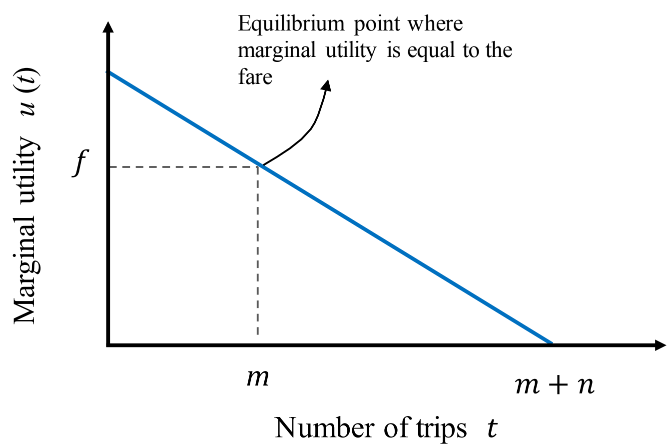


Figure 1: Marginal utility function.

2.2. The pass program

When the pass program is offered, riders can purchase a pass for a price of p [dollars] which grants access to unlimited free rides during the planning period. As pass owners pay a one-time fee (per planning period), they make as many trips as possible until they no longer obtain any positive marginal utility. Hence, the new equilibrium point occurs when $u(t) = 0$, which happens at $t = m + n$. Thus, pass owners have an economic incentive to complete all of their mandatory as well as non-mandatory trips. By completing their n non-mandatory trips, riders obtain an additional net utility equal to $nf/2$ which is depicted as a gray triangle in Fig. 2. Rational riders would only purchase a pass if their additional utility, $nf/2$, minus the cost of the pass, p, is larger than the cost they would incur if they did not purchase a pass (in which case they would pay mf [dollars] for m mandatory trips).

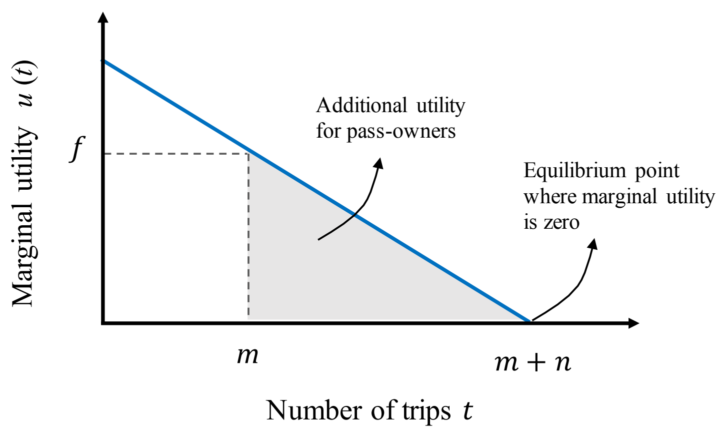


Figure 2: Marginal utility under the pass program.

2.3. The loyalty program

When the loyalty program is offered, riders that complete a total of l trips ("l" represents *loyalty trip threshold*), have to only pay αf [dollars] ($0 \leq \alpha \leq 1$) for the remainder of their trips in the planning horizon. The economic incentive of riders here is to complete their l trips so that they get a discount, α, on the rest of their trips. This loyalty program represents a single-tier structure (Kumar and Shah, 2004) where the threshold l is intuitively chosen to be larger than m. Otherwise, if $l \leq m$, riders get a discount on (mandatory) trips that they would have completed regardless of any loyalty program.

Riders who wish to enter the loyalty program first have to incur an additional cost of completing $l - m$ trips as shown in Fig. 3. When making each trip t, for $m < t \leq l$, the riders incur a cost (equal to f) that is higher than their marginal utility. This cost in total is equal to $(l - m)f - \int_m^l u(t)dt$ which is presented as a light-shaded triangle (upper-left triangle) in Fig. 3. After reaching the $(l)^{th}$ trip, the riders begin to see additional benefit from the loyalty program which is presented in Fig. 3 with a dark-shaded triangle (lower-right triangle). According to the geometry of the triangles in Fig. 3, the riders make a total of $t = m + n(1 - \alpha)$ trips where the marginal utility, $u\Big(m + n(1 - \alpha)\Big)$, is equal to the marginal cost, αf. The following lemma remarks the conditions required for a rider to enter the loyalty program.

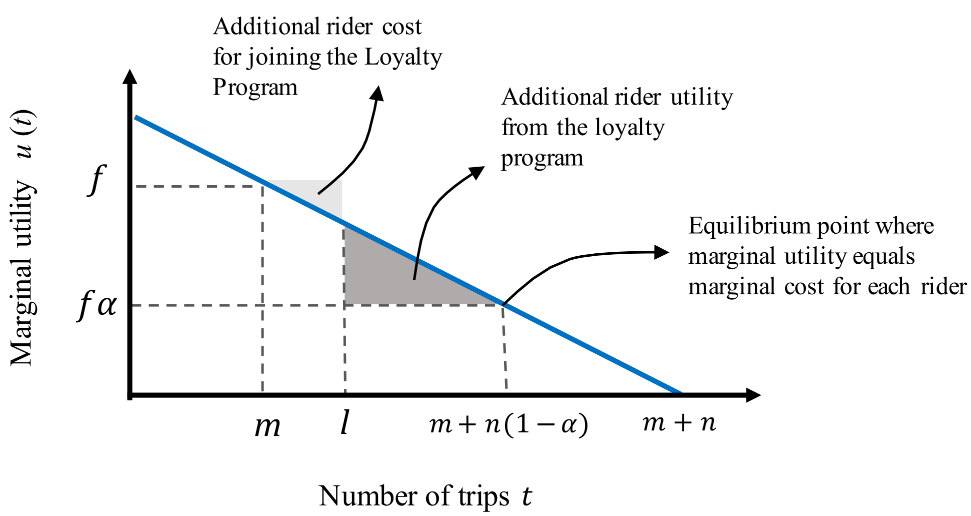


Figure 3: Marginal utility under the loyalty program.

Lemma 1. *Riders enter the loyalty program only if* $l \leq m + (1 - \alpha)n/2$.

Proof. For riders to enter the loyalty program, the dark triangle in Fig. 3 (i.e., additional rider utility) has to be larger than the lighter triangle (i.e., additional rider cost). Hence, a rider only enters the loyalty program if $f(l - m) - \int_m^l u(t)dt \leq \int_l^{m+n(1-\alpha)} u(t)dt - f\alpha\Big(m + n(1 - \alpha) - l\Big)$ which is equivalent to $l \leq m + (1 - \alpha)n/2$. □

References

Cervero, R., 1990. Transit pricing research. Transportation 17 (2), 117–139.

Chow, J. Y., 2014. Policy analysis of third party electronic coupons for public transit fares. Transportation Research Part A: Policy and Practice 66, 238–250.

Go Transit, 2016. Win it all back!
URL http://www.gotransit.com/public/en/fares/presto.aspx

Kumar, V., Shah, D., 2004. Building and sustaining profitable customer loyalty for the 21st century. Journal of retailing 80 (4), 317–329.

Lachapelle, U., Frank, L. D., 2009. Transit and health: mode of transport, employer-sponsored public transit pass programs, and physical activity. Journal of Public Health Policy 30 (1), S73–S94.

Li, T., van Heck, E., Vervest, P., 2009. Information capability and value creation strategy: advancing revenue management through mobile ticketing technologies. European Journal of Information Systems 18 (1), 38–51.

Pelletier, M.-P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies 19 (4), 557–568.

Pricing for a Last Mile Transportation System

Yiwei CHEN
Engineering Systems and Design, Singapore University of Technology and Design
8 Somapah Road, Singapore 487372
Tel: (65) 6876-4485; Fax: (65) 6779-5161; Email: yiwei_chen@sutd.edu.sg

Hai WANG, Corresponding author
School of Information Systems, Singapore Management University
80 Stamford Road, Singapore, 178902
Tel: (65) 9237-1568; Fax: (65) 6828-0919; Email: haiwang@smu.edu.sg

1. Problem Background

The *Last Mile Problem* (LMP)—that is, the design and provision of travel service from a public transportation node to a passenger's final destination—has attracted growing attention in recent years. The Last Mile Transportation System (LMTS) serves a public transportation node, such as a rapid transit metro station at which trains discharge passengers. Passengers' final destinations (homes, workplaces, public institutions, etc.) are spatially distributed in the urban area (or the last mile region) served by the node, and a fleet of vehicles is available to transport each passenger to her final destination. The routes and schedules of LMTS vehicles are flexible, and can adjust to specific last mile service requests.

Any passenger needing last mile service is required to provide advance notice to the LMTS of her impending arrival at the alighting station and her specific final destination. Once this information is received, the LMTS assigns her to one of the vehicles in the LMTS fleet, plans the vehicle's route so that it includes a stop at her destination, estimates the vehicle's departure time, and notifies her accordingly. Once all of the passengers assigned to a vehicle are on board, the vehicle executes a delivery route with stops at each passenger's destination and returns to the station to pick up passengers for its next delivery tour. Detailed LMTS settings for the area around the last mile region of one metro station can be found in Wang and Odoni (2014).

We study LMTS pricing with multi-type passengers—adults, senior citizens, children, and students. Given each type's last mile service demand in each last mile region, the geometric route configuration and corresponding vehicle's service travel time, discounts for specific passenger types, vehicle capacity, and cost, we propose and solve a constrained nonlinear optimization problem to determine the price for each passenger type and the service fleet size (number of vehicles) in each last mile region to maximize the social welfare generated by LMTS.

Our model is numerically implemented by using real data from Singapore. We show that by requiring the LMTS designer to offer discounted prices to special groups of passengers—senior citizens, children, and students—the optimal annual social welfare gained by implementing LMTS countrywide relative to the fraction of Singapore GDP contributed by the Singapore public land transportation service industry is 17.4%. We analyze a counterpart LMTS in which the LMTS designer sets the identical price for all types of passengers. We find that in the absence of price discounts for special groups of passengers, social welfare undergoes almost no change. Consumer surplus for LMTS passengers in special groups, however, suffers significantly.

2. Models

Consider a city that plans to implement an LMTS at a set of metro stations indexed by $1, \ldots, J$. There are I types of passengers, indexed by $1, \ldots, I$, who take trains and, potentially, use the LMTS. Trains dynamically arrive at station j over time. The inter-arrival time between two consecutive arriving trains is assumed to be a constant, denoted by h_j. Upon each train's arrival, N_i^j of type $i \in \{1, \ldots, I\}$ passengers are discharged. We assume that $N_1^j, \ldots, N_I^j$ are independent random variables. We denote by $\mu_{N_i^j}$ and $\sigma^2_{N_i^j}$ the mean and variance of N_i^j, respectively.

The LMTS designer aims to determine the price p_i for every type-i passenger who uses the LMTS. The price for each passenger type-i, p_i, is identical at all stations that are in the LMTS. The LMTS designer will also determine the capacity of vehicles used for last mile service, c, and the number of vehicles that provide LMTS at each station j, denoted by m^j. Each vehicle incurs per unit of time operating cost q. Service times for passengers who use the LMTS are independent and identically distributed. We denote by $\mu_{S_{\widehat{N}^j}}$ and $\sigma^2_{S_{\widehat{N}^j}}$ the expectation and variance of travel time, respectively, of one service trip (serving no more than c passengers) at station j if $\widehat{N}^j$ passengers are willing to use the LMTS from each train. At each station j, we denote by $\widehat{N}_i^j$ the number of type-i passengers who are willing to use the LMTS at price p_i. We denote by $\widehat{N}^j = \sum_{i=1}^{I} \widehat{N}_i^j$ the total number of passengers who are willing to use the LMTS. We denote by $\mu_{\widehat{N}^j}$ and $\sigma^2_{\widehat{N}^j}$ the mean and the variance of $\widehat{N}^j$, respectively.

Every passenger shares an LMTS vehicle with other passengers. A passenger who gets off the train and intends to use the LMTS is either (1) assigned to an idle vehicle that has an available seat or (2) directed to wait in a queue until a vehicle is available. Note that at each station j, the passenger expected waiting time for last mile service, w_j, depends on passenger destination topologies. Each passenger is endowed with a valuation (willingness-to-pay) for using the LMTS. For each passenger type i, passenger valuations are heterogeneous and supported on $[0, \bar{v}_i]$. The fraction of passengers whose valuations are no greater than v is denoted by $F_i(v) = v/\bar{v}_i$. We denote $f_i(v) \triangleq \mathrm{d}F_i(v)/\mathrm{dv}$ and $\bar{F}_i(v) \triangleq 1 - F_i(v)$.

Prices are determined in two steps. In the first, the LMTS designer determines price $p_i \in P$ for type-1 passengers as the system's benchmark price; the set P consists of a finite number of feasible prices. In the second step, to address fairness concerns, for each other passenger type $i \in \{2, \cdots, I\}$, the LMTS designer simply sets price p_i to be a fraction, $\theta_i \in [0, 1]$, of the system's benchmark price—i.e., $p_i = \theta_i p_1$ for all

$i \in \{2,\cdots,I\}$. Fairness fractions $\{\theta_i \in [0,1]: i = 2,\dots,I\}$ are determined before the LMTS designer assigns the benchmark price p_1. Vehicle capacity c is chosen from the set C that consists of a finite number of positive integers. Social welfare is the summation of consumer surplus and vehicle profit over all stations.

The LMTS designer determines the benchmark price p_1, vehicle capacity c, and number of operating vehicles m^j at each station j to maximize the LMTS's per unit of time expected social welfare $\mathrm{SW}(w_j, p_i)$. Therefore, the LMTS designer solves the following constrained nonlinear optimization problem:

$$\max_{p_1 \in P, c \in C, m^j \in \{0,1,..,M\}} \mathrm{SW}(w_j, p_i)$$

subject to: $$p_i = \theta_i p_1, \forall i \in \{2,\dots,I\} \tag{1}$$

$$w_j = g\left(\mu_{\widehat{N}^j}, \sigma^2_{\widehat{N}^j}; \text{destination features}\right), \forall j \in \{1,\dots,J\}. \tag{2}$$

$$\mu_{\widehat{N}^j} = \sum_{i=1}^{I} E\left(\widehat{N}_i^j\right) = \sum_{i=1}^{I} \mu_{N_i^j} \cdot \bar{F}_i\left(p_i + \alpha_i w_j\right). \tag{3}$$

$$\sigma^2_{\widehat{N}^j} = \sum_{i=1}^{I} Var\left(\widehat{N}_i^j\right) = \sum_{i=1}^{I} \sigma^2_{N_i^j} \cdot (\bar{F}_i\left(p_i + \alpha_i w_j\right))^2, \tag{4}$$

$$\rho_{queue\ utilization} = \frac{\mu_{\widehat{N}^j} \mu_{S_{\widehat{N}^j}}}{h_j m^j c} < 1, \forall j \in \{1,\dots,J\} \tag{5}$$

3. Numerical Experiments

We implement the pricing model described in our numerical experiments by using real data from Singapore. Singapore is a city-state with high utilization of public transportation, especially buses and metro services. We selected 10 sample metro stations which are located near residential areas well outside the downtown region. Distances between successive sample metro stations are longer than the distances between downtown stations, which causes the last segment of a passenger's trip to be longer; this renders LMTS more valuable—even necessary—in the last mile region around these sample stations. The three main types of metro cards are for adults, senior citizens, and children/students ($i = \{1,2,3\}$). All other parameters are selected based on real settings.

We carry out the following experiments and discussions.

1. *Impact of Prices on Social Welfare*

We explore the impact of LMTS price on social welfare. Given vehicle capacity c, for each possible LMTS price $p_1 \in P$ we compute the optimal number of vehicles m^j deployed to each station j that will achieve the optimal price-dependent expected social welfare, denoted as $\mathrm{SW}^*(p_1)$. Therefore, the optimal price dependent social welfare $\mathrm{SW}^*(p_1)$ is the optimal value of the optimization problem.

2. *Impact of Vehicle Capacity on Social Welfare*

The optimal price p^* declines as vehicle capacity increases. When a vehicle has more seats, due to economy of scale, the operating cost per passenger is reduced. In turn, the LMTS designer can charge passengers cheaper prices. Vehicle capacity c cannot be too small or too large. When c is too small, delivering service to passengers requires the LMTS to operate too many vehicles. Therefore, a large number of vehicles leads to prohibitively high operating costs. When c is too large, it is more likely that many seats on a vehicle for a given trip will be empty. Therefore, vehicle operating efficiency is too low. As a result, social welfare suffers when vehicle capacity is either too small or too large.

3. *Impact of Prices on Each Type of Passenger Consumer Surplus*

We analyze passenger behaviors and consumer surplus under the optimal LMTS price p^*. Roughly 90% of seniors, children, and students and half of the adults benefit from the LMTS. Therefore, our designed LMTS allows a majority of people—and, in particular, most of the people in special groups (senior citizens, children/students)—to benefit from the LMTS. The consumer surplus a senior, child, or student receives from using the LMTS is more than three times an adult's consumer surplus. Therefore, our designed LMTS can substantially improve people's welfare; in particular, such improvements are significant for senior citizens, children, and students.

4. *Necessity of Offering Discounts to Passengers in Special Groups*

We explain why offering discounts to special groups of passengers is necessary. We consider a counterpart LMTS with no discount for any passenger type, i.e., $\theta_1 = \theta_2 = \cdots = \theta_I = 1$. In this counterpart, the LMTS charges an identical price, denoted by p', which is the optimal price for all passengers. We report the optimal price in the counterpart LMTS with identical prices; the relative change in the percentage of each passenger type that uses the LMTS from the primary LMTS with type-dependent prices to the counterpart LMTS with identical prices. If the LMTS designer does not offer discounts to special groups of passengers, the decrement in the percentage of special groups of passengers who use the LMTS is much higher than the increment in the percentage of adults who are willing to use the LMTS. In addition, by eliminating discount offers, the percentage of consumer surplus loss for a special-group passenger who uses the LMTS is much higher than the percentage of consumer surplus gain for an adult who uses the LMTS.

5. *Impact of Total Social Welfare Generated by the LMTS*

We analyze social welfare. The optimal annual social welfare gained by implementing LMTS countrywide relative to the fraction of Singapore GDP contributed by the Singapore public land transportation service industry is 17.4%. This result demonstrates that an LMTS can play a significant role in improving living standards in Singapore. Our paper provides empirical support for the necessity of launching LMTS in Singapore.

References

Anderson, J. E. (1998). Control of personal rapid transit systems. Journal of advanced transportation, 32(1), 57-74.

Berger, T., Sallez, Y., Raileanu, S., Tahon, C., Trentesaux, D., & Borangiu, T. (2011). Personal Rapid Transit in an open-control framework. Computers & Industrial Engineering, 61(2), 300-312.

Bly, P. H., & Teychenne, R. (2005, May). Three financial and socio-economic assessments of a personal rapid transit system. In Proceedings of the tenth international conference on automated people movers (pp. 1-4).

Gómez-Ibañez, J., Tye, W., & Winston, C. (1999). Essays in Transportation Economics and Policy, 42. Brookings Institution Press, Washington D.C..

Lawphongpanich, S., & Yin, Y. (2012). Nonlinear pricing on transportation networks. Transportation Research Part C: Emerging Technologies, 20(1), 218-235.

Lees-Miller, J. D., Hammersley, J. C., & Davenport, N. (2009). Ride sharing in personal rapid transit capacity planning. Automated People Movers 2009, 321-332.

Liu, Z., Jia, X., & Cheng, W. (2012). Solving the last mile problem: Ensure the success of public bicycle system in Beijing. Procedia-Social and Behavioral Sciences, 43, 73-78.

Lou, Y., Yin, Y., & Lawphongpanich, S. (2010). Robust congestion pricing under boundedly rational user equilibrium. Transportation Research Part B: Methodological, 44(1), 15-28.

Lu, C. C., Mahmassani, H. S., & Zhou, X. (2008). A bi-criterion dynamic user equilibrium traffic assignment model and solution algorithm for evaluating dynamic road pricing strategies. Transportation Research Part C: Emerging Technologies, 16(4), 371-389.

Mueller, K., & Sgouridis, S. P. (2011). Simulation-based analysis of personal rapid transit systems: service and energy performance assessment of the Masdar City PRT case. Journal of Advanced Transportation, 45(4), 252-270.

Ng, K. (2015). Singapore feeling impact of rapidly ageing population. Today Online, Accessed Jul. 20, 2016. http://www.todayonline.com/singapore/singapore-feeling-impact-rapidly-ageing-population.

Singapore Average Monthly Wages, Trading Economics (2016), Accessed Jul. 20, 2016. http://www.tradingeconomics.com/singapore/wages.

Singapore Department of Statistics (2015). M601321 - Key Indicators By Detailed Industry In Transport & Storage Services, Annual. Accessed Jul.20, 2016. http://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=7147

Singapore Land Transport Authority (2012). Singapore Land Transport Statistics in Brief. Accessed Jul.20, 2016. https://www.lta.gov.sg/content/dam/ltaweb/corp/PublicationsResearch/files/FactsandFigures/ Statistics%20in%20Brief%202014.pdf.

Wang, H., (2016). Routing and Scheduling of a Last Mile Transportation System, working paper.

Wang, H., & Odoni, A. (2014). Approximating the Performance of a "Last Mile" Transportation System. Transportation Science, 50(2), 659-675.

Wu, D., Yin, Y., & Lawphongpanich, S. (2011). Pareto-improving congestion pricing on multimodal transportation networks. European Journal of Operational Research, 210(3), 660-669.

A Fare-reward Scheme for Commuters in Transit Bottleneck.

TANG, YILI [a], HAI, YANG [b]
[a] *Department of Civil and Environmental Engineering,*
The Hong Kong University of Science and Technology, China
Email: ytangap@connect.ust.hk
[b] *Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, China*
Email: cehyang@ust.hk

Abstract

This paper analyzes a new fare-reward scheme for managing commuter's departure time choice in a rail transit bottleneck, which aims to incentivize commuters' shift of departure time to the shoulder periods of the peak hour to relieve queueing congestion at the transit stations. A framework of the rail transit bottleneck is provided and the user equilibrium with a uniform-fare and social optimum with service run-dependent fares are determined. A fare-reward scheme (FRS) is then introduced that rewards a commuter one free trip during the shoulder periods after a certain number of paid trips during the peak hour. For a given number of peak-hour commuters and ex-ante uniform fare, the FRS determines the free fare intervals and the reward ratio (the number of paid trips required for one free trip, which is equivalent to the ratio of the number of rewarded commuters to the total number of commuters on each day during the peak hour). The new fare under the FRS is determined so that the transit operator's revenue keeps unchanged before and after introducing the FRS. Our study indicates that, depending on the original fare, FRS implements an optimal reward ratio up to 50% and yields a reduction of system total time cost and average equilibrium trip cost at least 25% and 20%.

Key words: Rail transit bottleneck, Queueing congestion, Fare-reward scheme, User equilibrium

1. Introduction

Since Vickery's bottleneck model (Vickrey, 1969), there has been a substantial stream of development of research in this area (Arnott et al., 1990; Laih, 2004; Lindsey et al., 2012). Unlike road traffic management, existing studies on transit bottleneck and queuing congestion are generally concerned with transit capacity choice, scheduling and fare pricing. Transit authorities usually adjust service frequency to accommodate variable passenger demand. Notably, applying the bottleneck model in mass transit, Kraus and Yoshida (2002) considered optimal fare and service frequency to minimize long term system costs. Various fare pricing schemes are considered on transit services in practice such as time-based differential pricing. For instance, the peak avoidance experiment in Netherland was implemented by rewarding commuters for travelling off-peak to reduce peak-hour ridership (Peer et al., 2016).

A major consideration of the above demand management is to offer monetary compensation or other incentives at the expenses of the government or the transit operator. With limited source of funding, governments or transit authorities would rather charge a high fare to passengers than avoid revenue losses. In addition, the above anonymous fare-free strategy also results in excessive off-peak ridership with majority coming from untargeted passengers who would otherwise travel by other modes.

To address the issues associated with the aforementioned demand management strategies for peak-hour transit operations; this paper introduces a controlled free fare reward scheme (FRS). Under the proposed FRS scheme, a commuter is rewarded one free trip during pre-specified shoulder periods after a certain number of paid trips during the peak hour. The planner can determine the rewarding ratio and periods to minimize commuters' trip costs and control the number of free commuters. In the meantime, it keeps the operator's revenue intact, which is in the same spirit of the 'revenue-neutral' traffic congestion management schemes in the literature, such as the Pareto-improving congestion pricing and revenue-refunding scheme (Guo and Yang, 2010) and the tradable credit scheme (Nie and Yin, 2013; Yang and Wang, 2011).

This work is organized as follows. First, we introduce the framework of urban rail transit bottleneck model with batch arrivals at work and the problems of user equilibrium with a uniform fare along with social optimum with service run-dependent fares. Second, the fare-reward scheme model is developed within the framework of the transit bottlenecks and seeks the optimal rewarding ratio and periods to minimize total commuting cost. Finally, we assess the system performances under the FRS in comparison with the original bottleneck situation. Sensitivity analysis of the results is conducted with respect to the initial system configuration and assumptions.

2. Theory development

2.1. Transit bottleneck and user equilibrium in the absence of FRS

Consider a single origin and destination connected by an urban rail line where each train is running with capacity s and headway h. Suppose a bottleneck occurs in peak hour, as shown in Fig. 1. During the peak hour, a given total number of N commuters travel with M uniformly service runs through the bottleneck. All trains through the bottleneck are indexed such that the service run 1 is the first train and M is the last train through bottleneck. Departure time of each service run is denoted by t_m, $m = 1, 2, ..., M$. The peak period lasts for $L = (M-1)h$.

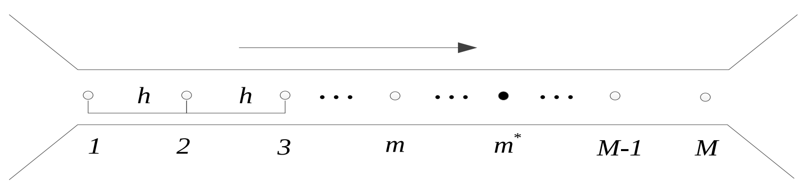


Fig. 1. The bottleneck of urban rail line

Kraus and Yoshida (2002) considered the optimal long-run peak-hour transit services without late arrival and showed that all service runs are full of capacity at social optimum. With reference to their results, the number of service runs is determined by

$$M = \frac{N}{s} \tag{1}$$

Equilibrium conditions require that commuters taking the first train and the last train do not incur queueing time and have identical trip costs consisting of only schedule delay costs and a uniform fare cost p_0 (hereinafter, '0' refers to the initial case). Namely,

$$\beta\left(m^* - 1\right)h + p_0 = \gamma\left(M - m^*\right)h + p_0 \tag{2}$$

The individual average equilibrium trip costs (AEC) is given by

$$\mathrm{AEC}_0 = \frac{\beta\gamma}{(\beta+\gamma)}(M-1)h + p_0 \tag{3}$$

and the system total time cost (TTC) is given by

$$\mathrm{TTC}_0 = \frac{\beta\gamma}{(\beta+\gamma)}(M-1)hN \tag{4}$$

2.2. The transit bottleneck model in the presence of FRS

Consider an original uniform fare p_0 throughout the whole period in the urban rail transit bottleneck. Implementation of the FRS changes the original uniform fare structure according to the commuters' entitlement of free ride. The peak hour of interest is divided into free fare interval (FFI) and uniform fare interval (UFI). UFI is the central period within the peak hour that spans an integer interval of runs $[i, j]$, including the preferred run m^* and with a uniform fare p, as shown in Fig. 2(a). FFI includes the two shoulder intervals before and after the UFI. After a certain number of paid rides, a commuter is entitled a free ride only during the FFI; a commuter without such an entitlement can choose either UFI or FFI at an uniform fare p, as shown in Fig. 2(b).

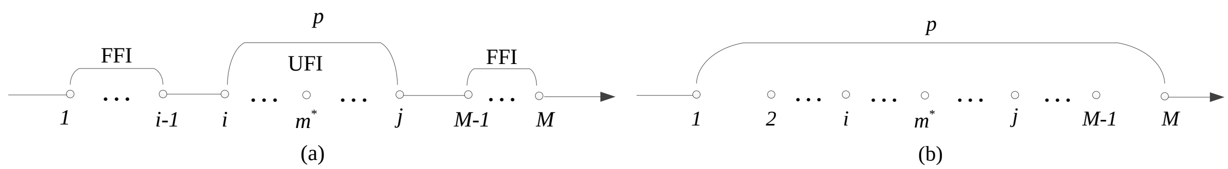


Fig. 2. Fare-reward scheme for (a) free ride commuters and (b) non-free ride commuters

The following assumptions are first introduced:

Assumption 1: The peak-hour operation of the rail transit system is under a long-term optimal configuration where all service runs are full of capacity.

Assumption 2: The total number of commuters and the rail transit headway remain the same and the operator's fare revenue keeps unchanged before and after implementation of the proposed FRS.

To ensure that a commuter prefers to take the free ride during FFI, the FRS must meet the following design criterion:

Design criterion 1: The FRS is designed so that the equilibrium trip cost in FFI with a reward (a free ride) is less than or equal to that in UFI without a reward.

Design criterion 2: The FRS chooses a reward ratio λ such that λM is an integer.

Under the above assumptions and criteria, travel assignment and fare structure for commuters is showed in Fig.3.

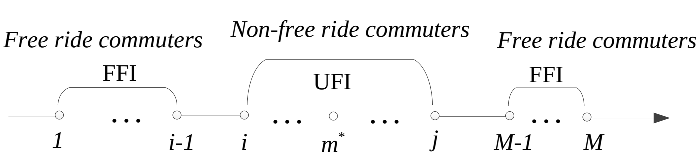


Fig. 3. Fare intervals under FRS with design criteria

Let the central UFI correspond to the service run interval $[i, j]$, as depicted in Fig. 3. From user equilibrium, $\beta(m^* - i)h + p = \gamma(j - m^*)h + p$, and design criterion 2, we readily have

$$\left(1-\lambda\right)M = j-i+1 \tag{5}$$

The total time cost of UFI commuters at equilibrium is

$$\mathrm{TTC}_{\mathrm{UFI}} = \frac{\beta\gamma}{\left(\beta+\gamma\right)}\left(j-i\right)h\cdot(1-\lambda)N = \frac{\beta\gamma}{\left(\beta+\gamma\right)}\left[\left(1-\lambda\right)M-1\right]h\cdot(1-\lambda)N \tag{6}$$

Similarly, the total time cost of FFI commuters at equilibrium is

$$\mathrm{TTC}_{\mathrm{FFI}} = \frac{\beta\gamma}{\left(\beta+\gamma\right)}\left(M-1\right)h\cdot\lambda N$$

For given M and p_0, the minimization problem of the system total time cost is given by

$$\min_{0\le\lambda\le1}\ \mathrm{TTC}(\lambda) = \theta\left(M-1\right)h\cdot\lambda N + \theta\left(M-\lambda M-1\right)h\cdot\left(1-\lambda\right)N \tag{7}$$

subject to

$$Np_0 = \left(1-\lambda\right)Np \tag{8}$$

$$\theta\left[\left(1-\lambda\right)M-1\right]h + p \ge \theta\left(M-1\right)h \tag{9}$$

where $\theta=\beta\gamma/\left(\beta+\gamma\right)$ is a constant. Constraint (8) is imposed to ensure the operator's revenue is unchanged before and after the FRS; inequality (9) is a mathematical statement of the aforementioned design criterion of the FRS. It is interesting to note that the FRS design problem reduces to a single variable optimization problem in terms of λ under a commute equilibrium constraint. In what follows, for convenience of mathematical analysis, the reward ratio λ is treated as a continuous variable for the optimization problem, $\lambda \in \left(0,1\right)$.

3. Performance of fare-reward scheme

One can analytically obtain the optimal solution of λ, as summarized in Table 1. As shown in Table 1 and Fig. 4 and Fig. 5, from the perspective of both system performance and individual commute cost, the FRS performs the best at a reference original fare $p_0^* = \theta Mh/4$, with an optimal reward ratio $\lambda^* = 1/2$. At this p_0^*, the absolute and percentage reduction in both system total time cost (TTC) and individual average equilibrium trip cost (AEC) reaches the maximum:

$$\Delta\mathrm{TTC}=\frac{\theta MhN}{4},\ \phi=\frac{M}{4M-4},\ \Delta\mathrm{AEC}=\frac{\theta Mh}{4},\ \varphi=\frac{M}{5M-4},\ \text{at}\ p_0^* = \frac{\theta Mh}{4} \tag{10}$$

Where ϕ is system efficiency defined as $\phi = \Delta\mathrm{TTC}/\mathrm{TTC}_0$, and φ is individual efficiency defined as $\varphi=\Delta\mathrm{AEC}/\mathrm{AEC}_0$.

In Table 1, Case (a), when $p_0 \le \theta Mh/4$. In this case, constraint (9) is always binding at optimum (commuter's equilibrium trip cost in FFI and UFI are identical), the feasible value of λ consists of two disjoint intervals belonging to $\left(0,1\right)$. For a given p_0, two solutions of λ exist to the system of simultaneous nonlinear equations (8) and (9). Moreover, at optimal solution λ^*, the trip cost of all commuters during the peak hour is equal to the right-hand side constant of constraint (9), which is also the time cost of free ride commuters in the two FFIs. The system total time cost, TTC, always decreases with p_0.

Case (b) when $p_0 > \theta Mh/4$. In this case, constraint (9) is always nonbinding, and any $\lambda \in (0,1)$ is feasible. The optimal λ^* is uniquely determined to be $1/2$, implying 'pay one get one free'. Different from Case (a), the minimum system total time cost, TTC, is a constant, independent of the original fare p_0. The average equilibrium commute cost (averaged over free and non-free rides), however, increases with p_0.

Finally, we point out that the average fare cost (for the same commuter over days or for all commuters on the same day) is always p_0 simply because the total number of commuters and the total revenue keeps unchanged before and after the FRS.

Table 1. Fare-reward scheme performances on urban transit rail bottleneck

p_0		λ^*	ΔTTC	ϕ	ΔAEC	φ
(a)	$0 \le p_0 \le \dfrac{\theta Mh}{4}$	$\dfrac{1}{2} \pm \sqrt{\dfrac{1}{4} - \dfrac{p_0}{\theta hM}}$	Np_0	$\dfrac{p_0}{\theta h(M-1)}$	p_0	$\dfrac{p_0}{\theta h(M-1) + p_0}$
(b)	$p_0 > \dfrac{\theta Mh}{4}$	$\dfrac{1}{2}$	$\dfrac{\theta MhN}{4}$	$\dfrac{M}{4(M-1)}$	$\dfrac{\theta Mh}{4}$	$\dfrac{\theta Mh/4}{\theta h(M-1) + p_0}$

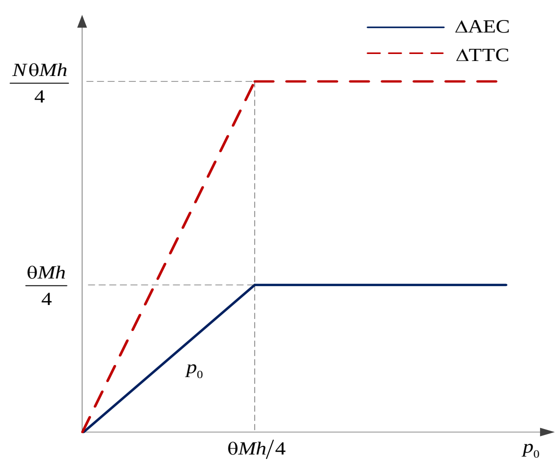


Fig. 4. The reduction in AEC and TTC

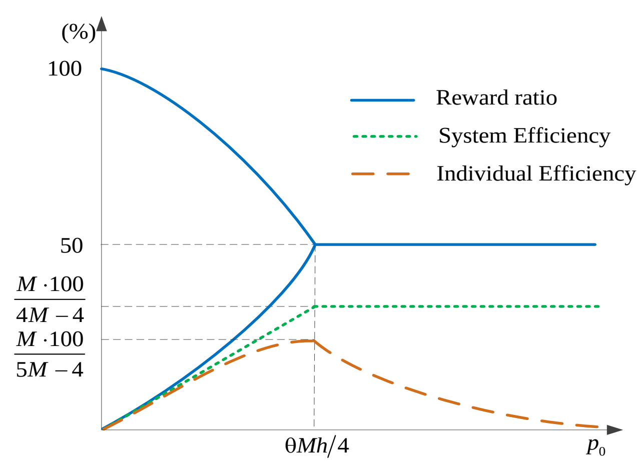


Fig. 5. λ and efficiency curves respect to original fare

4. Sensitivity analysis

The analyses of FRS so far are carried out for given supply (transit headway h and number of service runs M) and demand (number of commuters N and their shadow values of time β and γ). Here we conduct sensitivity analysis of the obtained results with respect to these exogenous inputs. For convenience, we fix the original fare at a reference point $p_0 = p_0^*$.

Fig. 6 plots the changes in system efficiency ϕ and individual efficiency φ with the total number of service run M, $M \in [2, \infty)$. It is observed that the system efficiency is always greater than the individual efficiency. In an extreme case with two service runs, system and individual efficiency reaches maximum values of 50.0% and 33.3%. When the number of service runs is sufficiently large, they approach their minimum value 25.0% and 20.0%, respectively. The positive minimum value of efficiency ensures the effectiveness of FRS in a busy rail transit line.

Fig.7 depicts the impact of the two exogenous parameters β and γ on system performances. Here we fix the shadow value of early arrival β, and look at the difference between the shadow value of late arrival and early arrival in terms of their ratio, γ/β. For a given combination (N,h) of demand and headway, ΔAEC is increasing with the value of γ/β over the range $\gamma/\beta \in (1,\infty)$. Particularly, ΔAEC reaches a minimum value $\beta Mh/8$ as $\gamma/\beta \to 1$ ($\theta \to \beta/2$) and a maximum value $\beta Mh/4$ as $\gamma/\beta \to \infty$ ($\theta \to \beta$) (late arrival is prohibited). These results illustrate that the FRS is more effective as late arrival penalty increases.

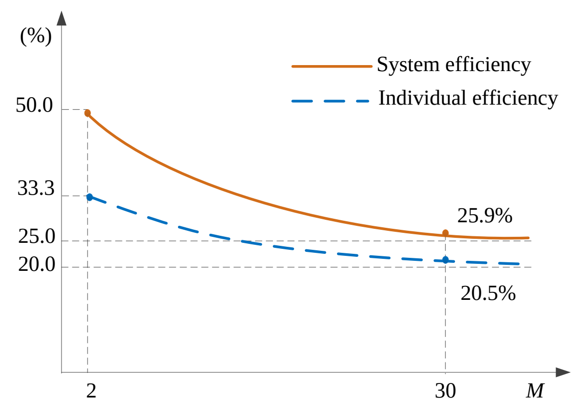


Fig. 6. Change of efficiency.

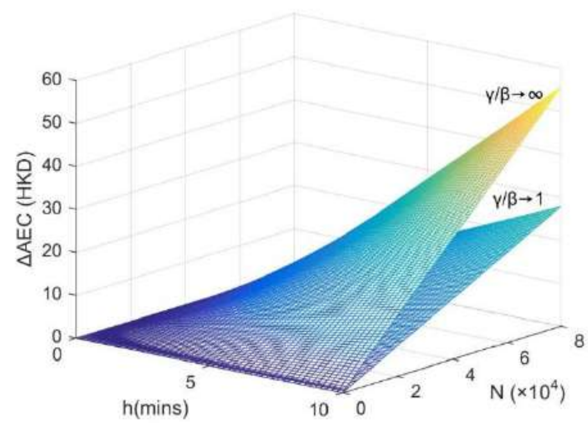


Fig.7. Change of ΔAEC

5. Conclusion

This paper proposed a novel fare-reward scheme (FRS) for managing peak-hour congestion of urban rail transit bottlenecks. It shifts commuters' departure time to reduce commuter queuing at stations in an incentive-compatible manner while keeping transit operator's revenue intact. Commuters' equilibrium choice of departure time is based on the trade-off between schedule delay and queueing time and their eligibility for a free ride. A free ride during the shoulder periods is granted after a certain number of paid trips during the peak hour.

We found that the performance of the FRS depends on the original fare. The best performance of the FRS is achieved at a reference original fare $p_0^* = \theta Mh/4$, at which 50% commuters shift departure time from the central to the shoulder period, giving rise to at least 25.0% and 20.0% reduction in total time cost and average equilibrium trip cost, respectively. The FRS is more effective as late arrival penalty becomes higher.

References

Arnott, R., De Palma, A., and Lindsey, R. (1990). Economics of a bottleneck. Journal of Urban Economics, 27(1), 111-130.

Guo, X., and Yang, H. (2010). Pareto-improving congestion pricing and revenue refunding with multiple user classes. Transportation Research Part B: Methodological, 44(8–9), 972-982.

Kraus, M., and Yoshida, Y. (2002). The commuter's time-of-use decision and optimal pricing and service in urban mass transit. Journal of Urban Economics, 51(1), 170-195.
Laih, C. (2004). Effects of the optimal step toll scheme on equilibrium commuter behaviour. Applied Economics, 36(1), 59-81.
Lindsey, R., van den Berg, V. A. C., and Verhoef, E. T. (2012). Step tolling with bottleneck queuing congestion. Journal of Urban Economics, 72(1), 46-59.
Nie, Y. M., and Yin, Y. (2013). Managing rush hour travel choices with tradable credit scheme. Transportation Research Part B: Methodological, 50, 1-19.
Peer, S., Knockaert, J., and Verhoef, E. T. (2016). Train commuters' scheduling preferences: Evidence from a large-scale peak avoidance experiment. Transportation Research Part B: Methodological, 83, 314-333.
Vickrey, W. S. (1969). Congestion theory and transport investment. The American Economic Review, , 251-260.
Yang, H., and Wang, X. (2011). Managing network mobility with tradable credits. Transportation Research Part B: Methodological, 45(3), 580-594.

Multimodal Transportation Services
FB1: Transit Networks
Friday 1:00 – 2:30 PM
Session Chair: Juan Carlos Munoz

1:00 An Internal Bounding Method for Line and Shuttle Bus Planning
*Evelien van der Hurk**
Management Science, DTU Management Engineering

1:30 Urban Transit Network Design and Timetabling Problem for Multi-Depot Round-Trip Routes
*James Chu**
National Taiwan University

2:00 The Limited-Stop Bus Service Design Problem with Stochastic Passenger Assignment
*Guillermo Soto, Homero Larrain, Juan Carlos Munoz**
Pontificia Universidad Católica de Chile

An Internal Bounding Method for Line and Shuttle Bus Planning

Evelien van der Hurk
Management Science, DTU Management Engineering
Technological University of Denmark
Email: evdh@dtu.dk

5 January 2016

1 Introduction

This work presents a new method for solving line planning, and specifically shuttle planning problems. The proposed Internal Bounding Method is an exact approach which, starting from a high level representation of the network, iteratively solves an extended representation of the network untill the optimal solution is found. In the worst case, the algorithm continues untill the network is represented in full detail. However the expectation is that for most practical cases convergence will be much faster, thus increasing computational speed and tractability of realisticly sized problems. The method will be applied to a case study for shuttle bus planning for the Danish railway, but the method is expected to also be applicable for general line planning problems.

2 Background

The shuttle planning problem is a restricted form of a line planning problem or network design problem. Schöbel (2011) presents an overview of methods and models in operations research for line planning. The shuttle planning problem was as one of the first studied by Kepaptsoglou and Karlaftis (2009). Jin et al. (2016) propose an optimization based approach for both the generation and selection of shuttle lines. van der Hurk et al. (2016) present an alternative method specifically focussed on the ability to include a large number of passenger groups, and realistic passenger route choice, in the optimization framework.

3 Problem Description

The shuttle bus design problem is to select the set of shuttle lines, at a specific frequency, in order to minimize passenger inconvenience given an operating budget. The below model is based on van der Hurk et al. (2016).

Given is a set of geographical lines (*geoline*) $\mathcal{G}$, each representing a possible sequence of stops a bus could serve. These lines consist of new shuttle bus services, regular bus services, slightly altered regular bus services to assist with the closure, and the rail lines. Each geoline g is associated with a set of frequencies $\mathcal{F}_g$. Each geoline and frequency has a minimum number of required vehicles ϕ_{gf}.

For every vehicle type $l \in \mathcal{L}$, δ_{lg} is the passenger capacity one vehicle on geoline g can provide per period, and k_{lg} represents the operating costs of a single vehicle of type l on geoline g. M_g represents the maximum number of passengers that can be transported on geoline g per period. β_l represents the maximum number of vehicles that are available of type $l \in \mathcal{L}$.

Furthermore a set of passenger groups $q \in \mathcal{Q}$ is defined, together with a set of routes per passenger group $\mathcal{P}_q$ that connect the passenger group's origin to the destination of a cost c_p through a set of trips connected in space and time. This cost reflects the waiting time, in-vehicle time, and transfer time on-route. Furthermore $\mathcal{P}(e)$ represents the set of paths that traverse a specific geoline-and-frequency trip. Each passenger group has a (demand) weight w_q.

Binary decision variables y_{gf} represent the decision to open a geoline g at frequency f. The route choice of passengers is reflected by the continuous decision variables x_{pq}, defining the number of passengers of group q selecting path p. Finally, integer decision variables v_{lg} represent the number of vehicles of type l assigned to geoline g, with $l \in \mathcal{L}_g$, the set of vehicle types that can be assigned to line g.

The shuttle bus design problem can be formulated similar to van der Hurk et al. (2016) as an integer linear program $1 - 10$:

$$\min \sum_{q \in \mathcal{Q}} \sum_{p \in \mathcal{P}_q} c_p x_{pq} + \sum_{g \in \mathcal{G}} \sum_{l \in \mathcal{L}_g} k_{lg} v_{lg} \tag{1}$$

subject to:

$$\sum_{p\in\mathcal{P}_q} x_{pq} = w_q \qquad \forall q \in \mathcal{Q} \tag{2}$$

$$\sum_{q\in\mathcal{Q}}\sum_{p\in\mathcal{P}_q(e)} x_{pq} \leq \sum_{l\in\mathcal{L}_g} \delta_{lg} v_{lg} \qquad \forall g \in \mathcal{G}, \forall e \in \mathcal{E}_g \tag{3}$$

$$\sum_{q\in\mathcal{Q}}\sum_{p\in\mathcal{P}_q(e)} x_{pq} \leq M_g y_{gf} \qquad \forall g \in \mathcal{G}, \forall f \in \mathcal{F}_g, \forall e \in \mathcal{E}_{gf} \tag{4}$$

$$\sum_{f\in\mathcal{F}_g} y_{gf} \leq 1 \qquad \forall g \in \mathcal{G} \tag{5}$$

$$\sum_{l\in\mathcal{L}_g} v_{lg} \geq y_{gf}\phi_{gf} \qquad \forall g \in \mathcal{G}, \forall f \in \mathcal{F}_g \tag{6}$$

$$\sum_{g\in\mathcal{G}} v_{lg} \leq \beta_l \qquad \forall l \in \mathcal{L} \tag{7}$$

$$x_{pq} \geq 0 \qquad \forall q \in \mathcal{Q}, \forall p \in \mathcal{P}_q \tag{8}$$

$$v_{lg} \geq 0 \qquad \forall l \in \mathcal{L}, \forall g \in \mathcal{G} \tag{9}$$

$$y_{gf} \in \{0,1\} \qquad \forall y_{gf} \in \mathcal{L} \tag{10}$$

The objective (1) minimizes expected passenger inconvenience and operating cost, which can be weighted by an appropriate scaling of c_p and k_{lg}. Constraints (2) impose the assignment of all passengers to a path. Constraints (3) restrict the assignment of passengers to paths to be within the available capacity of the geolines, at the selected frequencies, contained in the path. This capacity depends on the number, and the type, of the assigned vehicles. Constraints (4) ensures that passengers can only be assigned to a geoline and freqyency if that geoline is opened at that specific frequency. Constraint (5) imposes the selection of maximally one frequency f per geoline $g \in \mathcal{G}$. Constraints (6) require a minimum number of vehicles to be assigned to a geoline g at a specific frequency f. Note that the model allows to assign more vehicles than the minimum required to operate the line; which would in practice allow for a slightly higher frequency and thus a slightly lower travel time for passengers than estimated in the model. These situations will only occur when the additional assigned capacity is not sufficient to open the geoline g at the next higher frequency to the current selected f. Finally constraint (7) ensures that no more vehicles are assigned to a line than available.

4 Solution Method

One of the main advantages of the above formulation with a combined frequency and geoline decision variables, as first proposed in Claessens et al. (1998), is that it allows to include a minimal frequency constraint conditional on that the geoline is selected for operation. A main drawback is however that the number of decision variables grows rapidly in the number of frequencies included for consideration. This is mostly because the included paths per passenger group are frequency dependent.

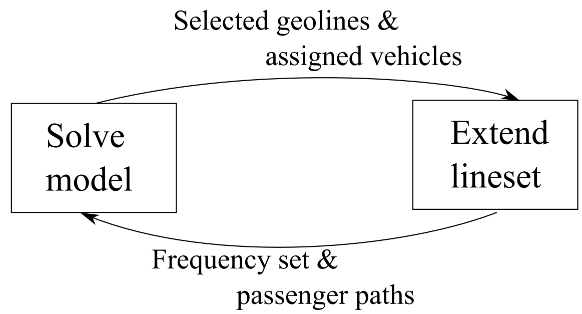


Figure 1: Internal Bounding Method

To increase tractability of solving this model, we propose a new *Internal Bounding Method*, as represented in Figure 1. It consists of two stages, which are solved iteratively. In the first stage, the model 1 – 10 is solved. However, not the full model is solved, but this step only considers a limited representation of the full network. Initially, only a single frequency is considered per geoline, which especially reduces the number of candidate paths per passenger group, each representing onde decision variable in the mode, substantially. Output of this step consists of the set of selected geolines and number of assigned vehicles to these lines. The latter gives an indication of the required capacity for the current passenger demand on the line.

In the second stage, this initial solution for the limited network is interpreted and new frequencies and paths are generated that seem likely to be attractive. These additional frequencies, and the additional paths, are used to extend the model 1 – 10, which is then resolved.

This algorithm is referred to as an *Internal Bounding Method* as each frequency and line represents not a feasible option, but rather represents a lowerbound on simultaneously the operator costs as well as the travel costs for the passengers. By extending the lineset, more variants are added for each line, thereby tightening the bounds, untill in the limit each possible line and frequency combination is represented. The algorithm stops as soon as a strictly feasible set of lines is selected, that is, the minimum capacity is assigned to each selected frequency of each selected geoline.

The hope and expectation is that only few line and frequency options are needed in the generation to find the optimal solution. The *Internal Bounding Method* is an exact approach that may also be of value to general lineplanning problems.

5 Outlook

The work in this project describes a new exact solution method for demand based design of shuttle bus services during track closures. The solution methods from this project are tested on data provided by Danish rail operator DSB and Danish bus operator Movia, and comparisons are drawn with their current approaches. The method is expected to increase the tractability of the model, thus enabling to solver larger and more realistic instances.

This research is part of the larger research project Integrated Public Transport Planning and Optimization (IPTOP) supported by both the IFD[1], and a consortium of Danish public transport operators such as train operator DSB and bus transport company Movia. IPTOP focuses on improving the coordination between separate transport providers to make the existing services more accessible, more reliable, and more attractive to travelers while operating efficiently at reasonable costs. For this it aims to develop new methods for global optimization of passenger preferences and operational constraints, leveraging today's better understanding of traveler data and preferences.

References

M. T. Claessens, N. M. van Dijk, and P. J. Zwaneveld. Cost optimal allocation of rail passenger lines. *European Journal of Operational Research*, 110:474–489, 1998.

Jian Gang Jin, Kwong Meng Teo, and Amedeo R. Odoni. Optimizing bus bridging services in response to disruptions of urban transit rail networks. *Transportation Science*, 50(3):790–804, 2016.

K. Kepaptsoglou and M. G. Karlaftis. The bus bridging problem in metro operations: conceptual framework, models and algorithms. *Public Transport*, 1(4):275–297, 2009.

Anita Schöbel. Line planning in public transportation: models and methods. *OR Spectrum*, 34(3):491–510, May 2011.

Evelien van der Hurk, Haris N. Koutsopoulos, Nigel Wilson, Leo G. Kroon, and Gbor Marti. Shuttle planning for link closures in urban public transport networks. *Transportation Science*, 50(3):947–965, 2016.

[1] Innovation Fund Denmark

Urban transit network design and timetabling problem for multi-depot round-trip routes

James C. Chu *

Abstract

This study solves the simultaneous planning problem of network design and timetabling for urban bus systems. Formulating and solving TNDTP are cumbersome because TNDTP combines all three major planning activities of TNP into a single problem. Therefore, the TNDTP considered in the literature is frequently size-limited and highly simplified (Guihaire and Hao, 2008). The first few models developed in related studies are solved in a sequential manner rather than as a single problem, e.g., the two-paper series of Shrivastav and Dhingra (2001) and Shrivastava and Dhingra (2002), as well as Quak (2003). Heuristics were developed for route construction in Shrivastav and Dhingra (2001), and the genetic algorithm was utilized for timetable synchronization in Shrivastava and Dhingra (2002). Quak (2003) adopted heuristic methods for routing and timetabling problems separately. The integration of network design and timetabling is not achieved in these studies. Zhao and Zeng (2008) proposed a simplified model for optimizing transit routes, headways, and timetables simultaneously. The model did not have a linear formulation, and thus, a method that combined simulated annealing, tabu search, and greedy algorithms was developed to solve the problem heuristically. For intercity bus systems, Yan and Chen (2002), Yan et al. (2006), and Yan and Tang (2008) also proposed integrated models for routing and timetabling, which were formulated as mixed-integer programming (MIP) models. One of the main differences between intercity and urban bus systems is that the former has greater freedom for timetabling, whereas the latter typically has specific restrictions for timetable patterns (Yan and Chen, 2002). For example, urban buses usually have constant headways because they are easy to remember and utilize for passengers. For intercity buses, passengers are accustomed to looking up the timetable before they travel, and thus, a rough service frequency will be sufficient. The second major difference is that the route design for intercity buses is flexible, whereas that for urban buses is not. For example, urban buses are generally required to return to the same depot for the convenience of vehicle management. On the contrary, this requirement does not always apply to intercity buses because their travel distances can be too long to

* *Corresponding author*, Associate Professor, Department of Civil Engineering, National Taiwan University, Taiwan, Tel: +886 2 33664235, Fax : +886 2 23639990, jameschu@ntu.edu.tw

enable them to return to the same depot in one day. Therefore, TNDSP for intercity bus systems is inappropriate for urban bus systems without significant modification. With regard to problem solving, the model in Yan and Chen (2002) was solved using an algorithm based on Lagrangian relaxation, whereas the models in Yan et al. (2006) and Yan and Tang (2008) were solved using heuristics algorithms. Yan et al. (2006) and Yan and Tang (2008) considered stochastic demand and stochastic travel time, respectively, and essentially belonged to a different class of problems. The preceding discussion indicates that the related TNDTP models are all solved using heuristic and relaxation approaches given the high complexity of the problem. Exact solution algorithms have not yet been successfully developed for TNDTP.

This study proposes an innovative model for urban TNDTP that focuses on urban bus systems with multi-depot round-trip routes. It is the first to propose a mixed-integer (linear) programming model for such a problem, which is the first major contribution of this work. To solve the aforementioned problem, a branch-and-price-and-cut (BPC) algorithm is developed. This algorithm is the first exact solution approach for such a problem, which is the second major contribution of this study. The proposed model includes multi-objectives for operator and passenger, unsatisfied demand, bus depot, route pattern, route length, frequency bound, headway structure, bus capacity, and transit assignment—all of which have never been considered in a single model in the literature. The model formulation for TNDTP and its solution algorithm have three noteworthy features. First, the transit routes are round-trip. In practice, round-trip routes are widely adopted in urban bus systems. They are attractive to both operators and passengers. For operators, round-trip buses return to the same depot after service, which simplifies the management of vehicles. Round-trip routes are also convenient for passengers because they can take the same route for both directions of their daily trips. Second, timetables can be either constant or variable headways, both of which are common practices in urban bus systems. A MIP formulation for constant headways is developed. Given this formulation, variable-headway timetables can also generated because they are simply relaxations of the constant-headway timetables. Third, the model formulation and solution algorithm apply the observation that a timetable of a bus line can be uniquely represented by its bus route and dispatch pattern. As illustrated in Fig. 1, the route for a bus line is the node sequence of 1-2-3-4-3-2-1, which forms a round-trip route. A complete timetable for the bus line can be derived without ambiguity given the dispatch pattern at the depot and the (deterministic) link travel times. MIP formulations that generate a timetable according to the bus route and dispatch pattern can then be developed based on this observation. The observation is also beneficial for the solution algorithm. Instead of pricing out a complete timetable during column generation, only a combination of bus route and dispatching pattern should be priced out, which significantly improves the efficiency of the solution algorithm.

A computational study is conducted to evaluate the performance of the proposed methodology and to understand the effects of model parameters on the results. The comparison of

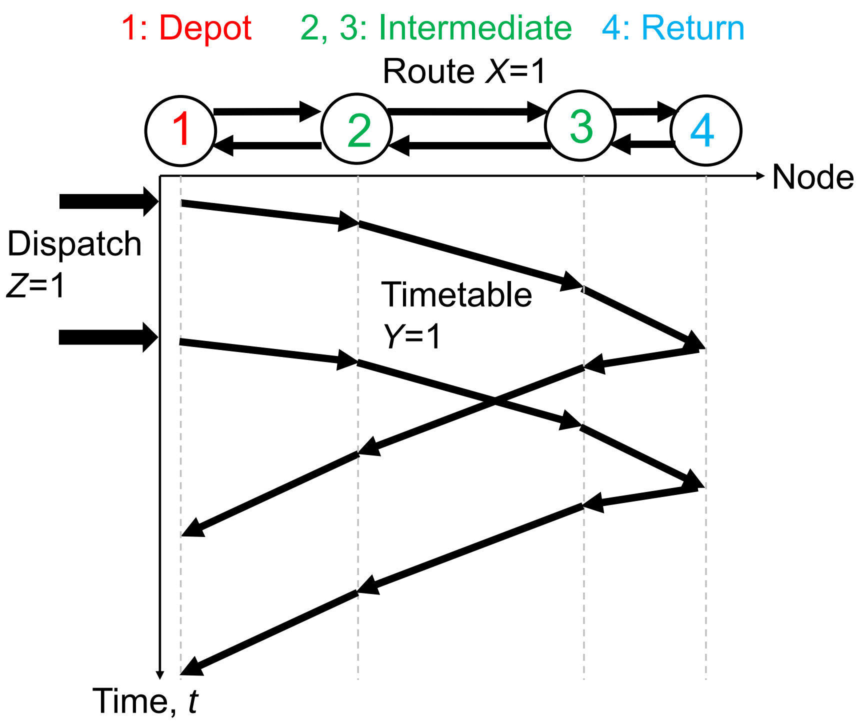


Figure 1: Concept of route, dispatching pattern, and timetable

the solution approaches using the same case clearly demonstrates that the proposed BPC algorithm is superior to MIP and the BP algorithm. Using BPC, a good quality solution can be found at the root node within a short period for most of the cases. Different values of objective weights, number of depots, number of candidate lines, minimum headway, and planning duration are tested. The outcomes of the routes and timetables as well as the statistics of operators and passengers are reasonable, which shows that the proposed model and the solution algorithm are useful tools for simultaneously planning bus network design and timetabling. By definition, the size of a time-expanded network increases with the number of periods. Consequently, problems formulated in a time-expanded network with a long duration are sometimes unsolvable. Benefiting from the Dantzig-Wolfe decomposition and the polynomial time algorithm for SP, the time for finding solutions does not increase drastically as the length of planning duration increases. This result indicates the potential for applying the methodology to real-world large problems.

Keywords: public transit, network design, timetable, branch-and-price-and-cut (BPC), mixed-integer programming (MIP), algorithm

References

Guihaire, V., Hao, J.-K., 2008. Transit network design and scheduling: A global review. Transportation Research Part A: Policy and Practice 42 (10), 1251–1273.

Quak, C., 2003. Bus line planning. Master's thesis, TU Delft.

Shrivastav, P., Dhingra, S., 2001. Development of feeder routes for suburban railway stations using heuristic approach. Journal of Transportation Engineering 127 (4), 334–341.

Shrivastava, P., Dhingra, S., 2002. Development of coordinated schedules using genetic algorithms. Journal of Transportation Engineering 128 (1), 89–96.

Yan, S., Chen, H.-L., 2002. A scheduling model and a solution algorithm for inter-city bus carriers. Transportation Research Part A: Policy and Practice 36 (9), 805–825.

Yan, S., Chi, C.-J., Tang, C.-H., 2006. Inter-city bus routing and timetable setting under stochastic demands. Transportation Research Part A: Policy and Practice 40 (7), 572–586.

Yan, S., Tang, C.-H., 2008. An integrated framework for intercity bus scheduling under stochastic bus travel times. Transportation Science 42 (3), 318–335.

Zhao, F., Zeng, X., 2008. Optimization of transit route network, vehicle headways and timetables for large-scale transit networks. European Journal of Operational Research 186 (2), 841–855.

The limited-stop bus service design problem with stochastic passenger assignment

Guillermo Soto
Department of Transport and Logistics Engineering
Pontificia Universidad Católica de Chile, Chile

Homero Larrain
Department of Transport and Logistics Engineering
Pontificia Universidad Católica de Chile, Chile

Juan Carlos Muñoz
Department of Transport and Logistics Engineering
Pontificia Universidad Católica de Chile, Chile

As demand and modal share of trips on public transport keeps rising throughout the world, the need for fast and reliable public transport systems with high quality standards for its users becomes more important. Bus Rapid Transit (BRT), which can be defined as a "high-quality, customer-orientated transit that delivers fast, comfortable and cost-effective urban mobility" (Wright, 2003), is a mass transportation alternative which has gained attention and popularity particularly during the last decade. Besides the emblematic cases of Curitiba and Bogotá, there are currently more than 200 cities around the world which have adopted BRT systems on their main arteries. Furthermore, new BRT corridors keep popping up in every corner of the world every day: nearly one third of these cities launched their BRT systems after the year 2010 (www.brtdata.org).

A key element in BRT allowing to provide fast rides while making an efficient use of a bus fleet, is the correct utilization of limited-stop services. Thus, counting on a reliable tool for designing efficient limited-stop services is of special importance in the light of the rise of BRT systems. During the last decade, many authors have proposed different methodologies for the design of limited-stop services (see [1], [2], [3], [4] and [5]). Different models and methodologies work under a wide range of assumptions, however, there is one particular assumption that all of these authors seem to share: the deterministic nature of passenger assignment.

Deterministic passenger assignment is a very convenient assumption which simplifies the formulation of the problem. In absence of capacity limitations (such as the bus capacity), it allows to formulate a mathematical programming problems that determines the optimal frequencies of a set of services while assigning passengers to the minimum cost route. However, this type of assignment (also known as “all or nothing”) has one important shortcoming: it assumes that every passenger will always take his/her shortest path, meaning that a slightly worse option will not carry any passengers at all. This triggers stability problems with the optimal solutions of the model.

To deal with this instability, we propose using a stochastic assignment. However, this poses some new challenges. Since there is no longer a natural way to solve the network design problem simultaneously, we separate the problem and solve it using the methodology suggested by [6].

We propose a framework for the limited-stop service design problem (LSDP) over a corridor where the problem is divided into the limited-stop service generation problem (LSGP), and the capacitated frequency optimization and passenger assignment problem (CFOAP).

The CFOAP can be presented in a generic way as the problem of minimizing a social cost function, subject to three groups of constraints: structural constraints, capacity constraints, and passenger behavior constraints. The input for this problem is a set of services previously defined by solving the LSGP or alternatively by an expert, an origin-destination (O-D) trip matrix, and some other key parameters.

In this framework, structural constraints encompass all the constraints that make the frequencies of a solution feasible: non-negativity and frequency conservation. The role of capacity constraints is to ensure that bus capacity (and possibly other types of capacity limitations) is not exceeded by the solution. The last group, passenger behavior constraints, guarantees that passengers are assigned to routes that are consistent to a selected behavioral model, and not just the routes that minimize social costs. In the deterministic case where passengers travel through their minimum cost option, behavioral constraints can be dropped only when i) there are no capacity constraints (or no capacity is binding) and ii) the user cost function in the total cost being minimized coincides with the cost individuals minimize to reach their destinations.

The solution of the CFOAP can tackled by first solving the uncapacitated version of the problem (FOAP) and then applying some heuristic approach to find a solution where capacity constraints are not violated, and where passenger behavior constraints are also met. Two greedy heuristics for capacity, which rely on solving the FOAP iteratively setting lower bounds to the frequencies, are proposed in [2] and [6]. In this work we solve the FOAP using a bi-level formulation that makes it possible to model passenger assignment as a stochastic process. The problem is divided into a frequency optimization stage (FOP) where passenger assignment is fixed, and a passenger assignment stage (AP) where frequencies are fixed. This separation of the problem in two levels makes it possible to test different behavioral models for passengers.

In this context the FOP consists on minimizing a social cost function subject only to structural constraints. This is a non-linear problem, but it is not hard to solve for instances inspired in real sized corridors. The AP, on the other hand, can be formulated as deterministic or stochastic. In this work, we consider that in the presence of parallel services passengers will choose a set of attractive services (which is a subset of the services connecting a given O-D pair) to perform their journey. It was proven by [7] that if a service A belongs to the set of attractive services for a desired trip, a service B which connects the same O-D pair with a lower travel time will also belong to the set of attractive services, regardless of its frequency. In simple words: if you are willing to take service A, but a faster service B shows up first, you will take it no matter how infrequent this service is.

To solve the deterministic AP, first we compute for each O-D pair the expected travel time associated to every possible set of attractive services. Then, we find the set of attractive services that minimizes this value using the methodology proposed by [7], and build an auxiliary network where every O-D pair is connected by an arc with travel time equal to this value plus a transfer cost. Finally, trips are assigned to their minimal route over this network, and passengers are split among the attractive services on every arc in proportion to the frequency of the services.

The stochastic AP, in turn, is solved by including randomness in two steps of the algorithm just described. First, for a given O-D pair, passengers will choose one of the reasonable sets of attractive lines following a multinomial logit model. The cost of an arc in the auxiliary network will be represented by the expected maximum utility of the pair. Then, we assume that passengers' route choice over this network can be also modeled as a multinomial logit. This last step of the assignment process can be performed using Dial's algorithm [8], which does not rely on route enumeration. The assigned trip over the network are finally translated into bus loads.

We have implemented both versions of the AP and solved the CFOAP for a test scenario with data inspired in a real corridor in the city of Santiago, Chile. Our results show, first of all, that the stochastic approach is more robust than the deterministic one, as expected. We performed sensitivity analysis on optimal solutions from the model showing that the predicted assignment from the deterministic approach has discontinuities that the stochastic approach avoid. These discontinuities in passenger assignment can be a serious issue when using the model to design a system, because the model may find a solution that satisfies capacity constraints only on a small neighborhood of the values of uncertain parameters such as the value of waiting time or operating costs.

Our tests have also shown that the bi-level approach permits solving larger instances in shorter times. This opens many new possibilities, especially when dealing with instances with constraining capacity. The heuristics to deal with capacity have to solve an instance of the FOAP in every iteration, which means that if the solution of this problem is more efficient, it is possible to design new capacity heuristics that explore a larger number of solutions, such as GRASP heuristics.

[1] Sun, C., W. Zhou, and Y. Wang. 2008. "Scheduling Combination and Headway Optimization of Bus Rapid Transit." Journal of Transportation Systems Engineering and Information Technology, 8(5), 61-67.

[2] Leiva, C., J. C. Muñoz, R. Giesen, and H. Larrain. 2010. "Design of Limited-Stop for an Urban Bus Corridor with Capacity Constraints." Transportation Research B, 44(10), 1186-1201.

[3] Chiraphadhanakul, V. and C. Barnhart. 2013. "Incremental Bus Service Design: Combining Limited-Stop and Local Bus Services." Public Transport, 5(1-2), 53-78.

[4] Larrain, H., J.C. Muñoz and R. Giesen. 2015. "Generation and Design Heuristics for Zonal Express Services." Transportation Research Part E, 79, 201 – 212.

[5] Chen, J., Liu, Z., Zhu, S., Wang, W. 2015. "Design of Limited-Stop Bus Service with Capacity Constraint and Stochastic Travel Time." Transportation Research Part E, 83, 1-15.

[6] Larrain, H. 2013. "Diseño de Servicios Expresos para Buses." PhD dissertation, Pontificia Universidad Católica de Chile.

[7] Chriqui, C. and P. Robillard. 1975. "Common bus lines". Transportation Science, 9, 115-121.

[8] Dial, R.B. 1971. "A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration." Transportation Research, 5, 83-111.

Multimodal Transportation Services

FC1: Containers Logistics

Friday 2:45 – 4:15 PM

Session Chair: Patrick Jaillet

2:45 Balancing the Trade-Off in Route Choice and Demurrage Costs in Inland Container Logistics
*Bernard Zweers**
CWI

3:15 An Integrated Model for Inbound Train Split and Container Loading in an Intermodal Railway Terminal
Bruno Bruck, Jean-François Cordeau, Emma Frejinger*
Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

3:45 The Stochastic Container Relocation Problem
[1]Virgile Galle, [1]Setareh Borjian Boroujeni, [2]Vahideh Manshadi, [3]Cynthia Barnhart, [4]Patrick Jaillet*
[1]MIT-Operations Research Center, [2]Yale School of Management, [3]MIT-Operations Research Center and Civil & Environmental Engineering, [4]MIT-Operations Research Center and Electrical Engineering & Computer Science

Extended abstract

Balancing the Trade-off in Route Choice and Demurrage Costs in Inland Container Logistics

Bernard Zweers
Center for Mathematics and Computer Science (CWI)
Amsterdam, the Netherlands

December 15, 2016

Abstract

A problem encountered in practice in inland container transportation is to decide on the mode and day of transportation of a container. To this end, we present an integer linear programming. The method presented leads to a significant reduction in cost compared to a method based on current practice.

Introduction

In container logistics, we encounter two major trends: (1) the increasing size of container vessels and (2) decreasing freight rates [3]. The first trend leads to an increasing number of containers that are delivered at one moment at a deep sea terminal. Besides the fact that this puts a large pressure on the terminal operations, the hinterland transport also needs to be optimized in order to guarantee an efficient dispatch of all containers. Together with the pressure to operate at the lowest cost, the use of a barge for the hinterland transport is encouraged, because a barge is cheaper than a truck and it can ship many more containers at once. The disadvantages of a barge are that a barge is slower than a truck and it is less flexible. In this extended abstract, a method to optimize the balance between transporting containers by truck and by barge is discussed.

We consider a commercial Third Party Logistics Provider (TPLP) in the Netherlands which needs to ship containers from multiple deep sea terminals in one port to one inland terminal. This planning for the containers is made per day. For each container one needs to decide on which day it is transported and which mode of transportation is used: truck or barge. As these are complex decisions, a model to optimize the decisions is needed.

All containers arrive by sea vessel at the deep sea terminal. Unloading a sea vessel takes, usually, 24 hours, so we assume that the first shipping moment of a container is one day after the arrival of the sea vessel. Moreover, it is assumed that when a container leaves the deep sea terminal by barge it arrives a day later at the inland terminal. When a container is shipped by truck, it arrives the same day at the inland terminal. Each customer has a latest delivery day, the so-called call date. The container must not be delivered after the call date. Since all customers are located in the neighborhood of the inland terminal, the container can arrive on the call date at the inland terminal. We are not focusing on the transportation from the inland terminal to the customer.

The moment a container arrives at the deep sea terminal, the demurrage period for that container starts. A container has a certain demurrage free period, if the container is still at the deep sea terminal after that period, one has to pay demurrage costs per day the container is at the deep sea terminal.

Each container has a certain size in Twenty foot Equivalent Unit (TEU) and a weight in kilograms. For each day it is known which ships are at the deep sea port and can be used to pick up containers. Each barge has a specific maximum capacity concerning both the total weight and the total TEU of the containers. Another constraint concerning barges is that a terminal might require a minimum number of containers to be loaded on a barge. Besides, each day there is an unlimited amount of trucks available. A truck can transport any container, so it has no constraints concerning weight or size. However, one truck can only ship one container.

The goal of our problem is to minimize the cost associated with the transportation and demurrage of containers and with the visit of a terminal by barge. The last one is added, because if a barge is visiting many terminals it might not have enough time to sail via all terminals. We do not decide on route the barge has to sail.

Surveys for planning problems for containers are given in [2] and [5]. Most research has been devoted to the strategic development of the network. If the operational aspect of a planning is considered, this is mostly done from the point of interest of a shipping company, which has usually no choice in the day of transportation. As pointed out in [5], most multiperiod operational planning is done for empty containers. The research that is done to the operational aspect of the planning of full containers [1, 4] does not differentiate between containers in their planning. So the problem discussed in this extended abstract is unique in its focus on a multiperiod planning problem for containers which all have different parameters, such as size, arriving date and call date.

Mathematical model

The mathematical formulation of the problem sketched in the introduction will be given in this section. Firstly, all input and decision variables will be explained and afterwards, an integer linear program (ILP) is given to solve the problem.

For every container $i = 1, \ldots, n_c$, we need the estimated arriving date (ETA) of the container at the deep sea terminal (da_i) and the call date of the customer(cd_i). Moreover, every container has a size in TEU (teu_i) and a weight in kilograms (w_i). Finally, for each container the deep sea terminal where it is located (tm_i) is known.

For each barge $b = 1, \ldots, n_b$, we know its capacity in both TEU (ct_b) and kilograms (cw_b). Moreover, the indicator po_{bt} indicates whether barge b is in the deep sea port in period $t = 1, \ldots, t_{max}$. Besides, for every terminal $r = 1, \ldots, n_r$ the minimum number of containers that needs to be loaded on barge if the terminal is visited by barge(mb_r) is given.

The costs associated with transporting container i in period t by truck is denoted as tr_{it} and by barge b as ba_{ibt}. The demurrage costs of leaving container i at the deep sea terminal for k days are given by dm_{ik}. Moreover, there is a penalty of tv_{rbt} for visiting terminal r by barge b on day t in order to reduce the number of terminals visited by barge. Another option to achieve that, could be to impose a constraint on the total number of terminals visited on one day by a barge. We have not chosen for this option, because it might occur that visiting one extra terminal can lead to an enormous decrease in transportation and demurrage costs. In this scenario, it might be better to visit that terminal and to have a delay in the barge schedule.

Finally, there are three types of binary decision variables. X_{ibt} indicates whether container i is loaded on barge b in period t. Y_{it} is the variable associated with the decision to transport container i by truck in period t. Finally, variable Z_{rbt} tells us whether terminal r is visited by barge b in period t.

Using the notation above, the problem can be formulated as the following ILP:

$$\min \sum_{i=1}^{n_c} \sum_{b=1}^{n_b} \sum_{t=1}^{t_{max}} \left(ba_{ibt} + dm_{i,t-da_i+1}\right) X_{ibt} + \sum_{i=1}^{n_c} \sum_{t=1}^{t_{max}} (tr_{it} + dm_{i,t-da_i+1}) Y_{it} + \sum_{r=1}^{n_r} \sum_{b=1}^{n_b} \sum_{t=1}^{t_{max}} tv_{rbt} Z_{rbt} \tag{1}$$

s.t.

$$\sum_{t=1}^{da_i} \left(\sum_{b=1}^{n_b} X_{ibt} \right) + Y_{it} = 0 \quad \forall i \tag{2}$$

$$\sum_{b=1}^{n_b} \sum_{s=da_i+1}^{cd_i-1} X_{ibs} + \sum_{t=da_i+1}^{cd_i} Y_{it} = 1 \quad \forall i \tag{3}$$

$$\sum_{b=1}^{n_b} \sum_{s=cd_i}^{t_{max}} X_{ibs} + \sum_{t=cd_i+1}^{t_{max}} Y_{it} = 0 \quad \forall i \tag{4}$$

$$\sum_{i=1}^{n_c} teu_i X_{ibt} \leq ct_b \quad \forall b, t \tag{5}$$

$$\sum_{i=1}^{n_c} w_i X_{ibt} \leq cw_b \quad \forall b, t \tag{6}$$

$$Z_{rbt} \leq po_{bt} \quad \forall r, b, t \tag{7}$$

$$\sum_{i:tm_i=r} X_{ibt} - Z_{rbt} mb_r \geq 0 \quad \forall r, b, t \tag{8}$$

$$\sum_{i:tm_i=r} X_{ibt} - Z_{rbt} n_c \leq 0 \quad \forall r, b, t \tag{9}$$

$$X_{ibt} \in \{0, 1\} \quad \forall i, b, t \tag{10}$$

$$Y_{it} \in \{0, 1\} \quad \forall i, t \tag{11}$$

$$Z_{rbt} \in \{0, 1\} \quad \forall r, b, t. \tag{12}$$

In the objective function (1), the transportation costs, demurrage costs and number of terminals visited by barge are minimized. In case the container is shipped before the end of the demurrage free period, the demurrage costs equal zero. Constraints (2) and (3) force together that the container is shipped a day after the ETA of the sea vessel and in time to arrive before or at the call date at the inland terminal. Equation (4) makes sure that the container cannot be shipped a second time from the deep sea terminal to the inland terminal. Without this constraint, this might happen at a certain terminal in order to reach the minimum number of containers to be picked up by barge. The two inequalities (5) and (6) ensure that the capacity of a barge is not exceeded. To prevent a terminal of being visit by a barge that is not at the deep sea port, constraint (7) is needed. The two inequalities (8) and (9) force, respectively, that when a terminal is visited the minimum number of containers is loaded and that no containers can be loaded on a barge without visiting the terminal by barge. Constraints (10), (11) and (12) ensure that the decision variables are binary. It is important to note that there always exists a feasible solution, because there are no restrictions on shipping a container by truck. Hence, it always feasible to ship all containers by truck.

Numerical results

Although the ILP formulation discussed in the previous section cannot deal with extremely large instances, it can be solved in reasonable time for an instance with 400 containers, 3 barges, 8 days and 10 terminals. The instance is based on a real data of a TPLP located in the Netherlands. The ETA of the containers is evenly distributed between $t = -1$ and $t = 6$ and the call date is between 3 and 10 days after the ETA. Each container has a size of 1, 2, or 3 TEU and a weight between 5,000 and 30,000 kilograms.

A barge schedule in which there is one barge available on every day, except at day 3 and 8, is given. Two of the barges have a capacity of 100 TEU and 1,500,000 kilograms and one barge of 125 TEU and 2,000,000 kilograms. We do not put any restrictions on the minimum number of containers to be loaded at a terminal, i.e., $mb_r = 1$. The costs are as follows: $tr_{it} = 150$, $ba_{ibt} = 25$ per TEU and the demurrage costs are 40 for each day after the demurrage free period for 1 TEU containers and 60 for containers of 2 or 3 TEU. We penalize visiting a terminal with $tv_{rbt} = 1$. In our scenario, the total barge capacity for the entire period is 650 TEU, whereas the total TEU of all the containers equals 665, so at least 15 TEU should be shipped by truck.

Table 1: Description of different scenarios

Scenario	Scenario Description
I	Basic scenario
II	ETA of container is 1 day later than in basic scenario
III	Call date of container is 1 day earlier than in basic scenario
IV	Capacities of the barges are 90% of the basic scenario
V	Barge of day 5 in basic scenario sails on day 3
VI	The demurrage free period is one day shorter than in basic scenario
VII	Demurrage costs are 50% of costs of the basic scenario
VIII	Trucking a container costs 75
IX	Visiting a terminal is penalized with $tv_{rbt} = 500$
X	Minimum number container loaded on barge per terminal equals 5

The data above sketches a basic scenario. We will slightly change some parameters of this basic scenario to obtain 9 other scenarios which are described in Table 1. In Table 2, the optimal costs of the ILP for these scenarios are compared with the solution of the first-in-first-out (FIFO) method, which is based on a method that is currently used in practice at the TPLP. In the FIFO method the jobs are ordered according to their ETA, call date and demurrage free days. Thereafter, the containers are planned in that order and each container is shipped at the earliest possible barge. The planning is made such that as many containers as possible are shipped by barge.

As can be seen in Table 2, the total cost of the optimal solution of the ILP is on average about 8% lower than the FIFO solution. This is mainly caused by the fact that, in general, the optimal ILP solution has lower demurrage costs. The ILP sometimes ships an container extra by truck in order to reduce the demurrage costs, which is never done by the FIFO method.

The scenarios II, III and IV are somehow easier than the basic scenario, because for each barge there are fewer containers that could be shipped with that barge. Hence, it is logical that for these scenarios the FIFO is performing better compared to the ILP solution than in the basic scenario. In scenarios V and VII, the demurrage costs are, relatively to the basic scenario, less important, so the FIFO method is in these scenarios also performing better than in the basic scenario. On the other hand, in scenarios VI and VIII, the demurrage costs are more important than in the basic scenario, so the ILP method is getting relatively better results. Finally, scenarios IX and X are the hardest scenarios, because the terminal where a container is located is a major issue. So it is not surprising that the FIFO solution is doing worse compared to the ILP solution in those scenarios than the basic scenario.

It should be noted that the total costs in Table 2 do not include the penalties for visiting terminals by barge. However, the ILP solution visits in almost all scenarios fewer terminals than the FIFO solution. In scenarios I up to VIII, 47-49 terminals are visited by barge in the ILP solution and 50-52 in the FIFO solution. In scenarios IX and X obviously fewer terminals are visited by barge. In scenario IX the ILP solution visits 23 terminals and the FIFO 26, but in scenario X the FIFO method visits fewer terminals than the ILP method, namely 35 versus 38. However, as can be seen in Table 2, this results in more than 4000 extra demurrage costs for the FIFO method.

Conclusion & future research

The model presented above can be used to decide which container should be transported on which day using which mode of transportation such that the total transportation costs, demurrage costs and terminals visited by barge are minimized. This leads to a substantial decrease in the transportation and demurrage costs and the number of terminals visited by barge compared to the FIFO method. Most of the decrease of the costs is caused by the fact that the ILP solution has fewer demurrage costs. The ILP method is better in making the trade-off between paying demurrage costs or transporting a container by truck. Hence, the more important demurrage costs are the better is the performance of the ILP method compared to the FIFO method. Moreover, if visiting a terminal by barge is a major issue, the ILP method is also a wise method to use.

Table 2: The results of the optimal solution of ILP compared with the First-In-First-Out method

	ILP			FIFO			
Scenario	Transport costs	Demurrage costs	Total costs	Transport costs	Demurrage costs	Total costs	Relative difference total costs
I	18025	960	18985	17600	3120	20720	8.4%
II	22200	820	23020	22200	1740	23940	3.8%
III	18425	1380	19805	17925	3420	21345	7.2%
IV	19850	780	20630	21050	780	21830	5.5%
V	17150	920	18070	17300	880	18180	0.6%
VI	19825	8380	28205	17450	15060	32510	13.2%
VII	17000	1100	18100	17450	1540	18990	4.7%
VIII	17050	780	17830	16775	3120	19895	10,4%
IX	19625	3280	22905	19125	6500	25625	10.6%
X	18400	1900	20300	17725	5960	23685	14.3%

Further research should be conducted to find a method which is able to solve larger instances of the ILP. A possibility might be to solve the linear relaxation and apply randomized rounding on the fractional solution. There lies a challenge in combining the rounding of both X_{ibt} and Z_{rbt} together. If we choose container i to be served by barge b on day t ($X_{ibt} = 1$), we need to visit terminal tm_i by barge b on day t ($Z_{tm_i bt} = 1$). However, it is not trivial that constraint (8) can be satisfied.

The current formulation can be extended to a situation in which there are multiple hubs in the transportation network, because the model decides for each container the day and mode of transportation on a certain route. In case a route might contain stops at multiple hubs, the only factor that makes the model more complicated is that consolidation constraints need to be added. Obviously, other modes than barge and truck can also be used in this model.

References

[1] Michael GH Bell, Xin Liu, Panagiotis Angeloudis, Achille Fonzone, and Solmaz Haji Hosseinloo. A frequency-based maritime container assignment model. *Transportation Research Part B*, 45:1152–1161, 2011.

[2] Marielle Christiansen, Kjetil Fagerholt, Bjorn Nygreen, and David Ronen. Ship routing and scheduling in the new millennium. *European Journal of Operational Research*, 228:467–483, 2013.

[3] Institute of Shipping Economics and Logistics. World Container and General Shipping. Technical Report 5/6, 2016.

[4] Dong-Ping Song and Jing-Xin Dong. Cargo routing and empty container repositioning in multiple shipping service routes. *Transportation Research Part B: Methodological*, 46(10):1556–1575, 2012.

[5] Nguyen Khoi Tran and Hans Dietrich Haasis. Literature survey of network optimization in container liner shipping. *Flexible Services and Manufacturing Journal*, 27(2-3):139–179, 2015.

An integrated model for inbound train split and container loading in an intermodal railway terminal

Bruno P. Bruck[a,b], Jean-François Cordeau[a,b], Emma Frejinger[a,c]

[a]*Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montréal, H3C 3J7, Canada*
[b]*HEC Montréal, Montréal, H3T 2A7, Canada*
[c]*Department of Computer Science and Operational Research, Université de Montréal, H3C 3J7, Canada*

1. Introduction

Intermodal freight transportation may be considered as one of the stepping stones of globalization, as it allows for efficient intercontinental door-to-door transportation of goods through a multimodal chain of land and sea transportation services that often involves several different carriers. In a classical example of an intermodal chain, loaded containers leave the initial shipper location by truck and are directed either to a port or to an intermodal terminal, from where a train will transport them to a port. A ship then moves the containers to another port, from where they are transported to the destination by one or a combination of several means of transportation (Crainic and Kim 2007).

These intermodal terminals are special transshipment nodes that are responsible for consolidating traffic and dispatching containers on trains destined to other nodes of the network, so that these containers can eventually reach their final destination. Although several studies in the literature concentrate on classification/shunting yards, they should not be confused with intermodal terminals, as only the latter allows temporarily storage, loading and off-loading of containers. For a comprehensive survey on the literature on intermodal transportation we refer the reader to the surveys of Crainic and Kim (2007) and of Steadieseifi et al. (2014).

Most of the intermodal traffic is containerized, because it ensures a safer, cheaper and more reliable means of handling the cargo. Indeed, this market has performed remarkably

Email addresses: `bruno.bruck@cirrelt.ca` (Bruno P. Bruck), `jean-francois.cordeau@hec.ca` (Jean-François Cordeau), `emma.frejinger@cirrelt.ca` (Emma Frejinger)

well in the last decade or so, with annual growth rates of about 15% (Steadieseifi et al. 2014). In North America alone, container traffic through ports has increased overall by 26.2% since 2010 (CBRE Research 2015).

While the international market mainly follows the ISO standard and uses 20, 40 and 45-foot containers, in North America there are also 53 and 48-foot containers, which are used for domestic traffic. Another complication of this market is that trains are usually double stacked and there are many types of railcars with different characteristics, e.g., number of *wells* (platforms), well length (40, 45, 48 and 53 feet long) and weight loading limit. This great variety of containers and railcar types has a significant impact on the *load planning*, which concerns the assignment of containers to slots of railcars. Performing a proper matching in the load planning is very important to ensure the best usage of the available capacity of railcars and also the fuel efficiency of the train, given that double stacked containers influence aerodynamic aspects.

The *block planning* is also another critical issue for the design of an efficient and profitable rail transportation system. A *block* is defined as a group of railcars, with possibly different origins and destinations, that are moved as a single unit between terminals. Because of this the railcars of the same block do not need to be handled individually at intermediate terminals, reducing handling costs.

In this paper we focus on an operational problem faced by a North American railway in the context of intermodal railway transportation of containers. On a daily basis, terminal operators have to take decisions concerning several different and interconnected activities, such as: how inbound trains are split into sequences of railcars and on which tracks those railcars are parked for loading and off-loading operations or even for temporarily parking; and the design of proper load and block plans. In our case, we assume that the block plan is partially given, as the expected total length of each block and the demand of certain types of railcars are known. However, the individual cars that compose each block is optimized based on the available resources. Then, instead of taking these aspects separately, we propose an integrated formulation and an iterative algorithm that incorporates all these decisions together in order to achieve better results.

2. Problem definition and solution method

We are given a sequence of inbound trains, a sequence of railcars arriving on these trains, a set of outbound blocks to be created, a set of outbound containers, and a set of tracks. Each inbound train can be split into a number of *segments*. A segment is composed by a set of railcars that occupy consecutive positions in the inbound train. Therefore, a segment can be defined by the first and last positions occupied by the railcars forming this segment. For example, a train comprising 100 railcars could be split into three segments: one from positions 1 to 35, one from 36 to 85 and one from 86 to 100. We assume that one can perform an *a priori* enumeration of the potential segments that can be created from any inbound train, and that the exact composition (in terms of railcars) and length of the segment is known with certainty. Each segment is further divided into a set of *sections* (or sub-segments) that will ultimately be assigned to different outbound blocks. For example, a segment with 35 railcars could be divided into a first section with 20 cars and a second one with 15 cars. The two sections will remain together until the departure of the outbound trains to which these sections are assigned.

The set of tracks can be partitioned into a set of storage and working tracks. While both types can be used to move railcars through the terminal and for temporary parking, containers can only be loaded on and off-loaded from railcars parked on working tracks. To model the fact that several segments can be parked on a single track at the same time, we discretize tracks into a number of *track slots* representing segments of a given length. For example, a 1000m track could be divided into 20 segments of 50m, thus a segment with length of 500m assigned to slot 1 would occupy the first 10 slots of this track. The sets of storage and working tracks can be further partitioned into a subset of single-ended tracks (where railcars enter and leave from just one end) and double-ended tracks (where railcars can enter and leave from both ends). In any case, the moves taking place at an open end must respect the last-in-first-out (LIFO) policy.

We assume that the inbound railcars belong to different families, which are defined based on well length. Then, there exists a demand (or minimum requirement) for railcars of certain types for each outbound block. This helps ensure that other terminals are supplied with the proper types and number of railcars. For instance, terminals on ports mainly have to deal with 20 and 40-foot containers and, thus, do not need railcars with wells longer than 40 feet.

The problem then consists in deciding on (i) how each inbound train is split, i.e., which segments are created; (ii) which track and track slot each segment is assigned to, (iii) which sections of each segment are assigned to each outbound block, and (iv) which containers are assigned to each railcar. These decisions are subjected to a large set of constraints that must be satisfied. In general terms, the main constraints are:

1) The set of segments created from an inbound train must be consistent, i.e., this set must form a partition of the railcars into subsets of consecutive elements.
2) Containers should be assigned to appropriate railcars. The assignment must take into

account not only the respective size (20, 40, 45, 48, or 53-foot) and type (e.g., high weight loading capability) of the container and railcar, but also additional constraints related to the loading of containers on railcars. In particular, a constraint is needed on the total number of containers assigned to the same railcar. Ideally, one would want to take into account all constraints of the associated load planning problem, but simplifications are necessary to keep the integrated problem tractable.

3) A container can only be assigned to a railcar if the railcar belongs to a section assigned to the appropriate block for the container.

4) The total number of segments created per train should respect the availability of the resources required for the switching.

5) The assignment of segments to tracks must respect the capacity of the tracks and ensure that segments assigned to the same track at the same time do not overlap (they must occupy disjoint positions). In addition containers can only be loaded on railcars that are parked on working tracks.

6) The sequence of moves (i.e., segments entering and sections leaving) on each track must satisfy the LIFO property given the planned arrival times of inbound trains and departure times of outbound trains. In the case of a double-ended track, care must be taken to properly represent the moves taking place at both ends of the track. A complicating factor that must also be taken into account is the presence of crossovers connecting different tracks.

7) Each outbound block has a certain demand for railcars of each family that should be met. In addition, these blocks have a certain expected length that should lie within a certain specified range.

8) The total length of an outbound train must respect the total capacity of the locomotives assigned to this train.

The objective function of the proposed *integer programming formulation* consists in minimizing the sum of two types of penalties. The first one concerns assignments of containers to railcars that are parked on a track slot far from the storage location of the container in the yard. This helps reduce the time required by cranes to transport containers from their location to the railcars. The second term refers to penalties applied to containers that are not assigned to any railcar. Note that this is necessary because it is not guaranteed that all containers can be loaded due to the temporal unavailability of space on the working tracks.

Instead of putting a heavy burden on the formulation in order to model the movements of cars in the terminal and ensure that they are feasible, we decided to relax the formulation and consider those aspects by adding lazy cuts to the model, in a *branch-and-cut* framework, for every new integer infeasible solution that is found during the branching process.

In practice, the problem can be solved on a daily basis for each terminal with the planned train schedule for the day. Indeed, in a first phase of the case study we tested the formulation on benchmark instances representing single days that were generated based on realistic data. These instances have about 3 to 4 inbound trains, an average of 250 inbound railcars (some of which are not loaded/off-loaded at the terminal) and 100 containers. The formulation was implemented in C++ using IBM Cplex 12.6.2. Preliminary tests showed that considering the complete enumeration of possible segments and sections for each train results in an incredibly high number of constraints and variables that render the formulation intractable. However, by considering a limited set of consistent segments for each train, we were able to solve the formulation in a few minutes.

Based on these observations we devised an iterative algorithm that solves the formulation at each iteration over a reduced set of segments. After each iteration the solution is analyzed and the best solution updated. Then, more segments are added and the formulation is solved once again, but using as a warm start the best solution found in previous iterations. This approach has showed some promising results and is being further improved.

References

CBRE Research. 2015 North America Ports and Logistics Annual Report, 2015. available at: http://www.cbre.us/research/2015-US-Reports/Pages/2015-North-America-Ports-Logistics-Annual-Report.aspx (Accessed on January 2, 2017).

T. G. Crainic and K. H. Kim. Intermodal transportation. In C. Barnhart and G. Laporte, editors, *Handbooks in Operations Research and Management Science*, volume 14, chapter 8, pages 467–537. Elsevier, 2007.

M. Steadieseifi, N. P. Dellaert, W. Nuijten, T. Van Woensel, and R. Raoufi. Multimodal freight transportation planning: A literature review. *European Journal of Operational Research*, 233(1):1–15, 2014.

The Stochastic Container Relocation Problem

Abstract

The growth of container shipping over the last decades significantly increased the interest in managing operations at container terminals more efficiently. These operations are commonly classified as: 1) *seaside operations*, involving the assignment of ships to quay cranes, the loading of export containers on vessels, or the discharging of import containers from vessels onto internal trucks; 2) *yard operations* including the routing of trucks within the yard, the stacking of containers for storage, or the delivery of import containers to trucks for delivery to another location. This work focuses on the latter problem.

Because storage space in yards is limited, containers must be placed on top of each other, thereby creating stacks. While this structure addresses space constraints, it complicates the process of delivering containers to trucks for delivery. Indeed, when a truck comes to pick up a container, the container can be blocked by other containers placed above it. In this case, blocking containers need to be relocated to other stacks in order to deliver the desired container. Consequently, many such relocation moves, also called reshuffles, are performed during the retrieval process of containers. These relocations are inherently inefficient and yards could incur considerable delays if they are not controlled. The *Container Relocation Problem (CRP)* (also known as the Block Relocation Problem) addresses this issue. It is concerned with finding a sequence of container moves that minimizes the number of relocations needed to retrieve all containers while respecting a given order of retrieval. Figure 1 provides a simple example of the CRP.

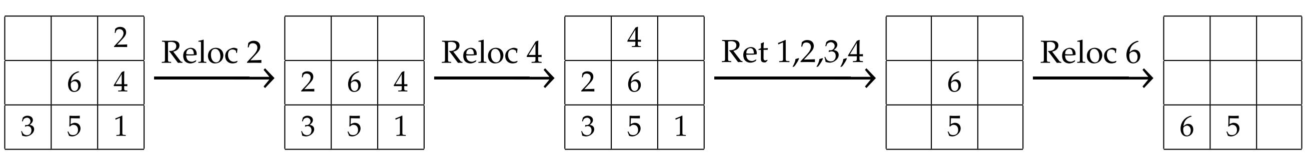


Figure 1: Configuration for the CRP with 3 tiers, 3 stacks and 6 containers. The optimal solution performs 3 relocations: relocate 2 on stack 1 on the top of 3; relocate 4 on stack 2 on the top of 6; retrieve 1; retrieve 2; retrieve 3; retrieve 4; relocate 6 on the empty stack 1; retrieve 5; finally, retrieve 6.

Researchers have tackled this particular problem from two point of views. The first approach is optimization-based and uses Integer Programming (Caserta et al. (2012), Petering and Hussein (2013), Zehendner et al. (2015)), branch and bound (Zehendner and Feillet (2014), Ünlüyurt and Aydın (2012), Expósito-Izquierdo et al. (2015)) or A^* (Zhu et al. (2012), Tanaka and Takii (2014), Borjian et al. (2015)). Because the problem is proven to be NP-hard by Caserta et al. (2012), the second approach is based on quick and efficient suboptimal heuristics, such as the ones presented in (Caserta et al., 2012), Wu and Ting (2010) or Wu and Ting (2012). In Kim and Hong (2006) and Zhu et al. (2012), lower bounds for the CRP are introduced. General review and classification surveys of the existing literature on the CRP and related problems can be found in Stahlbock and Voß (2008), Steenken et al. (2004) and Lehnfeld and Knust (2014).

One of the main assumptions of the CRP is full knowledge of the retrieval order of containers. However, because truck arrival times at the terminal are quite unpredictable due to uncertain conditions, the full information assumption is unrealistic. Nevertheless, new technology advancements such as Truck Appointment Systems (TAS) and GPS tracking can help predict information about relative truck arrival times: Hong Kong international Terminal (HIT) implemented the first TAS (1997) with 30-minute time slots, for which truck drivers can register (Murty et al. (2005)). Benefits of TAS are studied by Giuliano and O'Brien (2007) and Morais and Lord (2006). Recent information can be found in Phillips (2015) and Bonney (2015). One common type of information is a list of groups of containers likely to be retrieved within a given time window (time slots). Our study introduces the stochastic CRP (SCRP), which relaxes the full information assumption, and considers the same information scenario, *i.e.*, each container belongs to a group (representing a time slot), and retrieval orders between containers of the same group are equally likely.

Very few studies have tackled the SCRP, also referred to as CRP with Time Windows. In the original model of Zhao and Goodchild (2010), each container is assigned to a group, or "time window" such that all containers of a time window must be retrieved before any container of a later time window. Furthermore, the relative retrieval order of containers within a given time window is assumed to be a random permutation. For this model, Zhao and Goodchild (2010) develop a myopic heuristic (called RDH) and study, in different settings with two or multiple groups, the value of information using RDH. They conclude that a small improvement in the information system has a significant positive impact on the number of relocations. More recently, Ku and Arthanari (2016) use the same model as in Zhao and Goodchild (2010). After formulating the SCRP as a finite horizon dynamic programming problem, they suggest a decision tree scheme to solve it optimally. They also introduce a new heuristic called ERI (Expected Reshuffling Index), which outperforms RDH, and they perform computational experiments based on available test instances. Finally, there are two recent studies related to the SCRP, one on the Online Container Relocation Problem by Zehendner et al. (2016), and the second on an asymptotic average case analysis by Galle et al. (2016).

Our work starts by introducing a new way to model information, referred to as the *batch model*. This model applies in the case of high frequency retrievals of containers. The order of unknown containers is revealed by *batch*: for each group of containers, the full order of these containers is revealed and decisions to retrieve all these containers have to be made before any new information is revealed. This model differs from the one introduced by Ku and Arthanari (2016), in which containers are revealed one at a time (the *online* model). We derive a new family of lower bounds for which we show theoretical properties, and develop two new fast and efficient heuristics. Building on structural properties of the SCRP and taking advantage of the properties of the aforementioned lower bounds, we propose a novel optimal algorithm scheme based on decision trees and pruning strategies referred to as *Pruning-Best-First-Search* (*PBFS*). Due to the increasing complexity of the problem when time windows are "large", we introduce a second novel randomized algorithm referred to as *PBFSA* (PBFS-Approximate). It builds upon *PBFS* and Hoeffding's inequality to derive a sampling strategy resulting in an average error that we bound theoretically. Finally, we provide extensive computational experiments based on an existing set of instances. Various experiments are presented to show the efficiency of *PBFS* and *PBFSA* in different settings using existing instances presented in Ku and Arthanari (2016). The last experiment is used to conjecture the optimality of the leveling heuristic in the online model with no-information.

References

Bonney, J. (2015). Us ports move toward truck appointment model (04.24.2016). *Information Handling Services IHS*. Retrieved from http://www.joc.com/port-news/us-ports/port-new-york-and-new-jersey/us-ports-move-toward-truck-appointment-model_20150427.html.

Borjian, S., Galle, V., Manshadi, V. H., Barnhart, C., and Jaillet, P. (2015). Container relocation problem: Approximation, asymptotic, and incomplete information. *CoRR*, abs/1505.04229. http://arxiv.org/abs/1505.04229.

Caserta, M., Schwarze, S., and Voß, S. (2012). A mathematical formulation and complexity considerations for the blocks relocation problem. *European Journal of Operational Research*, 219(1):96–104.

Expósito-Izquierdo, C., Melián-Batista, B., and Moreno-Vega, J. M. (2015). An exact approach for the blocks relocation problem. *Expert Systems with Applications*, 42(17-18):6408 – 6422.

Galle, V., Borjian, S., Manshadi, V., Barnhart, C., and Jaillet, P. (2016). An average-case asymptotic analysis of the container relocation problem. *Operations Research Letters*, 44(6):723 – 728.

Giuliano, G. and O'Brien, T. (2007). Reducing port-related truck emissions: The terminal gate appointment system at the ports of los angeles and long beach. *Transportation Research Part D*, 12:460 – 473.

Kim, K. H. and Hong, G.-P. (2006). A heuristic rule for relocating blocks. *Computers & Operations Research*, 33(4):940–954.

Ku, D. and Arthanari, T. S. (2016). Container relocation problem with time windows for container departure. *European Journal of Operational Research*, 252(3):1031 – 1039.

Lehnfeld, J. and Knust, S. (2014). Loading, unloading and premarshalling of stacks in storage areas: Survey and classification. *European Journal of Operational Research*, 239(2):297–312.

Morais, P. and Lord, E. (2006). Terminal appointment system study. *Technical Report, Transportation Development Center of Transport Canada*. Retrieved from https://www.tc.gc.ca/media/documents/policy/14570e.pdf.

Murty, K. G., wah Wan, Y., Liu, J., Tseng, M. M., Lai, E. L. K.-K., and Chiu, H. W. (2005). Hongkong international terminals gains elastic capacity using a data-intensive decision-support system. *Interfaces*, 35(1):61 – 75.

Petering, M. E. and Hussein, M. I. (2013). A new mixed integer program and extended look-ahead heuristic algorithm for the block relocation problem. *European Journal of Operational Research*, 231(1):120–130.

Phillips, E. E. (2015). Southern california ports to try trucking appointment system (08.27.2015). *The Wall Street Journal*. Retrieved from `http://www.wsj.com/articles/southern-california-ports-to-try-trucking-appointment-system-1440711102`.

Stahlbock, R. and Voß, S. (2008). Operations research at container terminals: a literature update. *OR Spectrum*, 30(1):1–52.

Steenken, D., Voß, S., and Stahlbock, R. (2004). Container terminal operation and operations research a classification and literature review. *OR Spectrum*, 26:3 – 49.

Tanaka, S. and Takii, K. (2014). A faster branch-and-bound algorithm for the block relocation problem. *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, 14:7–12.

Ünlüyurt, T. and Aydın, C. (2012). Improved rehandling strategies for the container retrieval process. *Journal of Advanced Transportation*, 46(4):378–393.

Wu, K.-C. and Ting, C.-J. (2010). A beam search algorithm for minimizing reshuffle operations at container yards. *International Conference on Logistics and Maritime Systems, September*, pages 703 – 710.

Wu, K.-C. and Ting, C.-J. (2012). Heuristic approaches for minimizing reshuffle operations at container yard. *Proceedings of the Asia Pacific industrial engineering & management systems conference*, pages 1407 – 51.

Zehendner, E., Casserta, M., Feillet, D., Schwarze, S., and Voß, S. (2015). An improved mathematical formulation for the blocks relocation problem. *European Journal of Operational Research*, 245:415 – 422.

Zehendner, E. and Feillet, D. (2014). A branch and price approach for the container relocation problem. *International Journal of Production Research*, 52(24):7159 – 7176.

Zehendner, E., Feillet, D., and Jaillet, P. (2016). An algorithm with performance guarantee for the online container relocation problem. *European Journal of Operational Research*, 245:415 – 422.

Zhao, W. and Goodchild, A. V. (2010). The impact of truck arrival information on container terminal rehandling. *Transportation Research Part E: Logistics and Transportation Review*, 46(3):327–343.

Zhu, W., Qin, H., Lim, A., and Zhang, H. (2012). Iterative deepening A* algorithms for the container relocation problem. *IEEE Transactions on Automation Science and Engineering*, 9(4):710–722.

Multimodal Transportation Services

FD1: Maritime Shipping and Fleets

Friday 4:30 – 6:00 PM

Session Chair: Stein W. Wallace

4:30 Speed Optimization Across Different Emission Control Zones

[1]Line Reinhardt, [2]Christos Kontovas*
[1]Aalborg University, [2]Department of Maritime and Mechanical Engineering-Liverpool John Moores University

5:00 A Column-Row-Generation Approach to Liner Shipping Network Design

*Jun Xia, Zhou Xu**
Hong Kong Polytechnic University

5:30 Planning for Charters: A Stochastic Maritime Fleet Composition and Deployment Problem

[1]Xin Wang, [1]Kjetil Fagerholt, [2]Stein W. Wallace*
[1]Norwegian University of Science and Technology, [2]Norwegian School of Economics

Speed optimization across different emission control zones.

L. Reinhardt and C. Kontovas

Maritime transport is the backbone of international trade and a key engine driving globalization. In addition to being efficient from an economic perspective, the global maritime chain has to significantly improve its environmental image (Psaraftis and Kontovas [5]). Air pollution from ships such as NO_x and SO_x is currently at the center stage of discussion by the world shipping community and measures with the aim at reducing the environmental externalities of maritime transport will get increased attention.The introduction of emission control areas (ECAs) is one of the air emissions related regulations faced by the industry and more regulations are planned in the future, see Figure 1. However it is still uncertain how these regulations affect the maritime companies and their operations. According to Stopford [6], the bunker cost is 35% to 50% of operational cost of a vessel and according to Mærsk Line [1] around 21% of the company expenses in 2013. The recent focus on emissions results is yet another problem for management to tackle in order to ensure efficient maritime operations under these emission regulations.

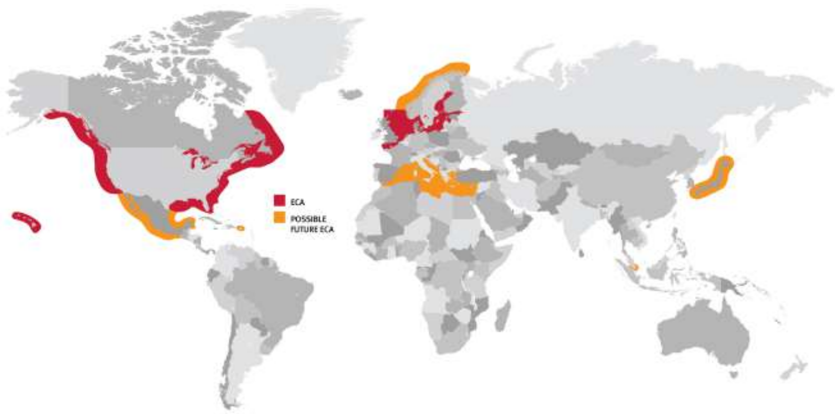


Figure 1: Current and planned emission control areas.

The latest review on the maritime routing, see Christiansen et al. [2], highlights the fact that mainly due to increasing price of bunker fuel, more attention has been devoted to sailing speeds and the environmental impact of ships. This is also in line with the survey by Wang et al. [7] who note that minimization of the environmental impact is becoming more important in designing shipping networks and they analyze recent papers that deal with the optimization of vessel sailing speed as a measure to reduce both operating costs and emissions. A recent paper by Fagerholt et al. [3] optimizes the bunker cost and traversal point when traversing an emission control area showing that the optimal speed and traversal point changes to ensure a lower overall cost.

In this presentation we apply the concepts presented by Fagerholt et al. [3] to the area of liner shipping considering scheduling and transit times as well. The emissions of CO_2 and SO_X under the cost minimized operations satisfying emission regulations are evaluated to see if the regulations may have undesired side effects. The same ideas are used to evaluate the cost and emission consequences of proposed policies on liner shipping operations. These Models can also be used to compare different policies.

1 Problem and Model

The problem presented here is to evaluate the effects of the emission regulations on the sailing speeds of the vessels and to evaluate the emissions of the vessel when minimizing bunker cost alone compared with minimizing bunker cost combined with external costs. The work is primarily relevant for vessels sailing crossing the ECA zones borders. In the speed optimization models we include the transit restrictions.

1.1 Model

The model presented below is related to a liner shipping company that seeks to minimize costs by optimising speed under the constrains of predefined transit times.

type	notation	description
sets	L	set of legs where $l \in L$ is a leg on the service.
	Q	set of demands defined by a route (as set of legs $l \in L$) between two ports A to B.
	P_l	set of secants used for approximating the bunker curve on leg $l \in L$
parameters	W_r	number of weeks used for the round trip of a service
	F_w	the frequency of the vessels in hours. Here it is always 168 hours (weekly).
	P_l	the port stay of leg $l \in L$
	T_l^m	minimum time used for sailing leg $l \in L$ (at maximum speed)
	H_q	transit time requirement of demand $q \in Q$
	f_l	has value 1 if leg l is a start leg of a service and zero otherwise.
	ϕ_l^p	gradient of secant $p \in P$ of leg $l \in L$
	ω_l^p	y-axis intersection of secant $p \in P$ of leg $l \in L$
variables	t_l^d	(continuous) departure time of leg $l \in L$ at its end port
	t_l^a	(continuous) arrival time of leg $l \in L$ at its end port
	C_l	(continuous) cost of sailing leg $l \in L$

Table 1: Overview of notation used in the model

An overview of the notation can be seen in Table 1. Moreover we use the notation $(l', l) \in L$ to indicate that l' is the previous leg of l. Legs $l \in L$ that are partly in different emission zones are separated in to a leg for each emission zone.

$$M1: \qquad min \sum_{l \in L_{eca}} D_{0.1\%} C_l + \sum_{l \in L \setminus L_{eca}} D_{380} C_l \tag{1}$$

$$t_l^a - t_{l'}^d + F_w W_r f_l \geq T_l^m, \qquad \forall r \in R,\ (l', l) \in L_r \tag{2}$$

$$\phi_l^p (t_l^a - t_{l'}^d + F_w S_r f_l) + \omega_l^p \leq C_l, \qquad \forall r \in R,\ (l', l) \in L_r,\ p \in P_l \tag{3}$$

$$\sum_{(l,l') \in L_q} \Big(t_l^d - t_{l'}^d + F_w S_r f_l\Big) - P_q^s \leq H_q, \qquad \forall q \in Q \tag{4}$$

$$t_l^a, t_l^d \geq 0,\ C_l \geq 0, \qquad \forall r \in R, l \in L_r \tag{5}$$

The objective (6) minimizes the sum of bunker cost where $D_{0.1\%}$ is the bunker price for fuel, BW0.1%, used in ECA and D_{380} is the bunker price for the fuel, BW380, which contains 3.5% Sulphur. For the bunker prices we use the Bunkerworld 380 Index (BW380) and Bunkerworld 180 Index (BW0.1%) which are unweighted simple averages of all BBP prices for IFO380 and IFO 180 respectively. The values used for $D_{0.1\%}$ and D_{380} can be seen in Table 3. The parameter C_l is the bunker cost on leg l. Constraints (2) ensure that the legs are not traversed at a faster speed than the maximum speed of the vessel. For every leg the consumption is restricted by a set of linear functions represented by constraints (3). In constraints (3) the variable ϕ_l^p is the slope of secant p_l on leg $l \in L$ and ω_l^p is the intersection of the secant. The transit time is ensured to be below the requirement H_q for demand $q \in Q$ with constraint (4).

Pollutant	$/Ton	parameter
CO_2	37	E_{CO_2}
SO_x	12700	E_{SO_x}

Table 2: External costs of emission per ton of CO_2 and SO_x presented in [4].

Bunker type	CO_2 (Ton)	SO_x (Ton)	price $
BW380	3.114	0.07	370
BW0.1%	3.206	0.002	620
Parameter	V_{380}, $V_{0.1\%}$	S_{380}, $S_{0.1\%}$	D_{380}, $D_{0.1\%}$

Table 3: Emission of CO_2 and SO_x of the two bunker types BW380 used in areas with no regulations and BW0.1% used in the Baltic emission control zone.

Clearly this model finds the optimal speeds on the legs so that the fuel cost is minimized under the regulations imposed by the emission control zones. However minimizing the cost of the bunker may not result in the best emission profile.

To investigate the sulphur and CO_2 emission profile of the solution we look at the external costs of CO_2 and SO_x. Table 3 shows the amount of CO_2 and SO_x emitted when using the two bunker types BW380 and BW0.1% which are respectively used outside and inside the emission control zone.

The information from Table 2 and 3 is used for a new objective which includes the external costs of CO_2 and SO_x ensuring that the cost of society is considered along side the bunker cost when scheduling a path of a vessel.

$$M2: \quad min \sum_{l \in L_{eca}} (D_{0.1\%} + V_{0.1\%} E_{CO_2} + S_{0.1\%} E_{SO_x}) C_l + \sum_{l \in L \setminus L_{eca}} (D_{380} + V_{380} E_{CO_2} + S_{380} E_{SO_x}) C_l \quad (6)$$

The set $L_{eca} \subset L$ is the subset which only contains the legs which are inside the ECA zone.

2 Results

We have tested the model and the two objectives on a test based on the service presented in Figure 2. In the test instance two transit time constraint to be satisfied are introduced.

Figure 2: Liner shipping service used for preliminary tests.

The preliminary results show that we get a

In Table 4 the costs of the solutions achieved by using model $M1$ and $M2$ are shown. For $M1$ we have calculated the external costs of the emission for the solution which minimizes the

Test	Bunker cost	Emission cost	total cost	Avg Speed		vessels
				BW380	BW0.1%	
M1	838815	432774	1271559	16.1	13.7	3
M2	876078	322133	1198211	13.0	15.8	3

Table 4: Preliminary test results of running M1 and M2 on the test instance.

Zone	M1, Ton		M2, Ton	
	CO_2	SO_x	CO_2	SO_x
BW380	1154	25.946	750	16.852
BW0.1%	1348	0.841	1789	1.116
Total	2502	26.787	2539	17.968

Table 5: Test results showing the total CO_2 and SO_x emissions for the solutions to M1 and M2.

bunker cost. It is clear that when only looking at bunker costs then the solution of $M1$ is 4.35% less expensive that that of $M2$ however the external costs of the solution of $M1$ is 29.31% larger than that achieved by the $M2$ model. Considering both external costs and bunker costs then the solution of $M2$ is 5.94% cheaper than that of $M1$.

3 Conclusion

The preliminary results show that although emission control areas may help to reduce the emission inside them, companies have to increase the speeds, and thus emissions, outside the control areas. In October 2016, IMO decided that by 2020 the global sulfur cap in bunkers should be dropped to 0.5%. Additional tests should, thus, be performed on future scenarios. In addition, future work includes the investigation of route planning under these sulfur requirements.

References

[1] Annual report. A.P. Møller - Mærsk A/S., 2013. Copenhagen, Denmark.

[2] M. Christiansen, K. Fagerholt, B. Nygreen, and D. Ronen. Ship routing and scheduling in the new millennium. *European Journal of Operational Research*, 228(3):467 – 483, 2013.

[3] K. Fagerholt, N. Gausel, J. Rakke, and H. Psaraftis. Maritime routing and speed optimization with emission control areas. *Transportation Research Part C: Emerging Technologies*, 52:57–73, 2015.

[4] A. Korzhenevych, N. Dehnen, J. Bröcker, M. Holtkamp, H. Meier, G. Gibson, A. Varma, and V. Cox. Update of the handbook on external costs of transport, final report for the european commission. European Commission, RICARDO-AEA, Oxfordshire, UK, 2014.

[5] H. Psaraftis and C. Kontovas. Balancing the economic and environmental performance of maritime transportation. *Transportation Research Part D*, 15(8):458–462, 2015.

[6] M. Stopford. *Maritime Economics, 3rd ed. Table 13.9.* Routledge, Oxford, UK, 2009.

[7] S. Wang, Q. Meng, and Z. Liu. Bunker consumption optimization methods in shipping: A critical review and extensions. *Transportation Research Part E: Logistics and Transportation Review*, 53(1):49–62, 2013.

A Column-Row-Generation Approach to Liner Shipping Network Design

Jun Xia and Zhou Xu*
Department of Logistics and Maritime Studies
The Hong Kong Polytechnic University, Hong Kong

1 Introduction

In liner shipping, container ships rotate among seaports to transport cargos with a regular service frequency [6]. The sequences of port calls constitute a service network on which carriers operate their daily transportation services. Due to the capricious nature of the shipping industry, carriers constantly need to adjust their service networks to maintain their competitiveness in response to an ever-changing market. Liner Shipping Network Design (LSND) aims at creating a set of regular services (or rotations) for a designated fleet of oceangoing ships to transport containerized cargos. Containers can be transshipped from one ship to another at an intermediate port in order to improve a carrier's transportation efficiency, as well as to extend its market coverage. The major objective of LSND is to maximize the carrier's total profit, this being the total revenue from satisfied demands minus the total operating cost, which includes any transshipment costs.

LSND has been widely investigated in recent literature (see [7] for a comprehensive review). It is well known that optimizing LSND, even with zero transshipment cost, is computationally challenging, as it is at least as hard as solving a set-covering problem, which is known to be strongly NP-hard [4]. Without transshipment costs, LSND can be seen as a variant of service network design with asset management (SNDAM), where multiple types of assets are to be deployed on cycles to maintain certain frequencies of designated services. Many existing models for SNDAM, such as those in [2, 3, 5], are derived based on a multiple commodity flow (MCF) network, in which each arc has a capacity that is aggregated among its passing cycles so as to impose a limit on the flow of its carrying cargo. Due to the decomposable structure of these models, the column generation approach can often be applied in solving their linear programming (LP) relaxation effectively, which provides bounds on optimal objective values that are useful in promoting the developments of exact methods for solving for the problem. However, in these studies, the cost of transshipment is ignored, mainly due to the hardness of capturing exact amounts of transshipped cargos in these MCF based models.

There are only a few existing models that have taken into account transshipment costs for LSND [1, 11, 4, 9, 8]. From these models, it can be seen that including the transshipment costs has complicated the mathematical formulations of LSND, as both the number of decision variables and the number of constraints are proportional to the number of all feasible rotations, which can be exponentially many. Therefore, even solving a linear programming relaxation of these models becomes very challenging, causing the traditional column generation approach to be no longer applicable. As a result, most existing works focus only on finding heuristic solutions, and are unable to report any optimality gaps that can be used to measure the quality of their obtained solutions.

*Corresponding email: lgtzx@polyu.edu.hk

The contribution of this work is threefold. First, we propose for LSND a new mixed-integer linear programming (MILP) model where transshipment costs are well captured. Since, in this new model, both decision variables (columns) and constraints (rows) are also proportional to the number of feasible rotations, there can be a large number of columns and rows, which makes the model very challenging to solve. Second, in this work we have developed a new optimization method, referred to as a Column-Row-Generation (CRG) approach, to solve the LP relaxation of the new model, which provides an upper bound on the optimal solution for LSND. Third, we have embedded this CRG approach into a branch-and-price framework to compute optimal or near-optimal solutions for LSND.

2 Problem Formulation

We formulate the optimization problem of LSND based on a newly defined planning network that consists of voyage nodes, transshipment nodes, voyage arcs, and transshipment arcs. Figure 1 illustrates such a network with two rotations. Port calls in different rotations are differentiated, and cargos on the transshipment arcs are used to represent the transshipment costs in the objective function. Ships sail at the designed speed, and all services need to maintain a weekly frequency.

We use a vector of integer variables $\mathbf{y} \in \mathcal{Y}$ to determine the rotations that are selected to be operated, use a vector of continuous variables $\mathbf{x} \in \mathcal{X}$ to determine the cargo demands that are fulfilled, and use a vector of continuous variables $\mathbf{f} \in \mathcal{F}$ to represent the cargo flows carried by each operated rotation. For $\mathbf{f}$, it must satisfy flow balance constraints on the voyage nodes of each operated rotation. For $\mathbf{f}$ and $\mathbf{x}$, they together need to satisfy flow balance constraints on the transshipment nodes. For $\mathbf{x}$, it needs to satisfy that the fulfilled demand for each origin-destination pair cannot exceed the maximum available amount. For $\mathbf{y}$, it must satisfy that the number of deployed ships cannot exceed the available fleet size. For $\mathbf{f}$ and $\mathbf{y}$, they together need to satisfy that for each rotation, the cargo flows on each voyage arc of the rotation cannot exceed the total capacity of ships deployed to the rotation. Accordingly, let $\mathcal{C}_1$ and $\mathcal{C}_2$ denote the flow balance constraints on voyage nodes and transshipment nodes. Let $\mathcal{C}_3$, $\mathcal{C}_4$ and $\mathcal{C}_5$ denote the demand, fleet and capacity constraints, respectively. Note that constraints $\mathcal{C}_1$ and $\mathcal{C}_5$ are defined for each rotation. Given a revenue vector $\mathbf{p}$, a cargo based cost vector $\mathbf{c}$, and a rotation based cost vector $\mathbf{a}$, we can obtain the following MILP

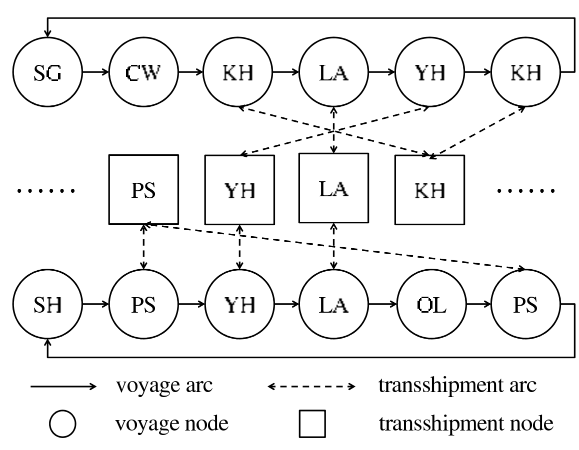


Figure 1: An example of the planning network with two rotations

model for the optimization problem of LSND:

$$\begin{aligned} \text{(P)} \quad & \max \quad \mathbf{px} - \mathbf{cf} - \mathbf{ay} \\ & \quad s.t. \quad \mathbf{f} \in \mathcal{C}_1; \quad (\mathbf{f}, \mathbf{x}) \in \mathcal{C}_2; \quad \mathbf{x} \in \mathcal{C}_3; \quad \mathbf{y} \in \mathcal{C}_4; \quad (\mathbf{f}, \mathbf{y}) \in \mathcal{C}_5. \end{aligned}$$

3 Solution Method

In model (P), both the total number of decision variables and the total number of constraints grow with the total number of feasible rotations. It is therefore not affordable to directly solve the model using a general integer programming solver, especially when a large number of feasible rotations have to be taken into account. Moreover, since there are constraints of (P) that are defined each rotation, such as constraints $\mathcal{C}_1$ and $\mathcal{C}_5$, the traditional column generation approach is not applicable to solve the LP relaxation of the model.

In this work, we extend the column generation approach to a novel CRG approach, so as to solve the LP relaxation of model (P) effectively. Different from many other row (or cut) generation techniques that produce valid inequalities to strengthen the relaxations of integer programs (see [10, 12] for example), the rows generated in our CRG approach interrelate with the rotations, which are necessary to validate the formulation of model (P). The idea behind the CRG approach is to work on a restricted master problem (RMP) that consists of only variables and constraints that are associated with a subset of rotations. New rotations are generated iteratively to induce a simultaneous generation of new columns and rows added to the restricted formulation. The CRG approach continues to generate and add new columns and rows until no new rotations can be found to improve the optimal objective value of the RMP. Compared with the traditional column generation approach, the most critical step of the new CRG approach is the construction of values of the missing dual variables when some constraints of $\mathcal{C}_1$ and $\mathcal{C}_5$ are absent in the RMP. To tackle this, we have proposed several effective construction methods. Based on their constructed values, new rotations can be efficiently evaluated and generated. When the CRG approach stops iterations, it can be proved that the optimal objective value of the RMP equals the optimal objective value of the LP relaxation of model (P), which provides a valid upper bound for model (P). Based on this, we embed the CRG approach into a branch-and-price framework that can find optimal or near-optimal solutions to model (P).

4 Computational Results

First, in order to compare the performances of our new CRG approach and the general MIP solver CPLEX, we have randomly generated test instances based on the data of Baltic (12 ports, and 22 origin-and-destination pairs for demands) and the data of WAF (20 ports, and 37 origin-and-destination pairs for demands) in a benchmark suite provided by [4]. These instances have different numbers of ports (ranging from 6 to 18) and different numbers of origin-destination pairs for demands (ranging from 6 to 34). Each random instance is represented by "pX-dY-Z", implying that the instance involves "X" ports and "Y" origin-destination pairs for demands, and is generated from data "Z", in which the data of Baltic and the data of WAF are denoted by "Z=b" and "Z=w", respectively.

To apply CPLEX to solving model (P), we have to generate all feasible rotations in advance. We implement the branch-and-price algorithm with our CRG approach embedded by C. We set the time limit as two hours for both CPLEX and the branch-and-price algorithm. The results are reported in Table 1, where columns "OBJ" present the objective values of solutions obtained, columns "UB" present the upper bound values, columns "GAP" present the optimality gaps, columns "TIME"

Table 1: Comparison with CPLEX for solving instances that are randomly generated.

		CPLEX				Branch-and-price			
INST	COL	OBJ	UB	GAP(%)	TIME	OBJ	UB	GAP(%)	TIME
p6-d6-b	732	102797	102797	0.0	13	102797	102797	0.0	1
p6-d6-w	395	684748	684748	0.0	55	684748	684748	0.0	3
p6-d10-b	511	277019	277019	0.0	27	277019	277019	0.0	3
p6-d10-w	467	1074382	1074382	0.0	2277	1074382	1074382	0.0	2889
p9-d12-b	4190	368314	476473	22.7	7200	374040	374040	0.0	3
p9-d12-w	4702	1172185	1619487	27.6	7200	1281405	1281405	0.0	2631
p9-d15-b	4393	612629	943813	35.1	7200	712921	712921	0.0	61
p9-d15-w	4702	2996155	3590359	16.6	7200	3202709	3247785	1.4	7200
p12-d20-w	901184	n/a	n/a	n/a	n/a	3632189	3878197	6.3	7200
p12-d22-w	893776	n/a	n/a	n/a	n/a	2710615	2906484	6.7	7200
p15-d26-w	957809	n/a	n/a	n/a	n/a	3734023	4127781	9.5	7200
p15-d28-w	909432	n/a	n/a	n/a	n/a	5121447	5553856	7.8	7200
p18-d32-w	$> 10^9$	n/a	n/a	n/a	n/a	5579441	6240528	10.6	7200
p18-d34-w	$> 10^9$	n/a	n/a	n/a	n/a	5405048	6175337	12.5	7200

present the computation times, and column "COL" presents the number of all feasible rotations that need to be generated for CPLEX.

From Table 1, it can be seen that when all feasible rotations are generated in advance, CPLEX can, within the time limit, solve the instances with 6 ports to optimality, and achieve an optimality gap of 16.6% $\sim$ 35.1% for the instances with 9 ports. Since the number of feasible rotations increases exponentially with the network size, CPLEX fails to output any feasible solution for larger instances within the time limit, simply reporting "n/a". In comparison, *our branch-and-price algorithm solves all the instances with 6 ports and most of the instances with 9 ports to optimality within 1 hour, significantly outperforming CPLEX in terms of both solution quality and time efficiency.* For other instances, as the size of the test instances increases, our branch-and-price algorithm can produce near-optimal solutions with optimality gaps varying from 1.4% to 12.5%.

Next, to compare the performances of our new CRG approach and the existing solution methods proposed in [9] (denoted by Plum) and in [4] (denoted by Brouer), we use six original benchmark instances of the data of Baltic and WAF, these being indicated by p12-d22-b-l, p12-d22-b-b, p12-d22-b-h, p20-d37-w-l, p20-d37-w-b, and p20-d37-w-h, respectively. For these instances, the results for Plum's method are available in [9]. Note that the results reported in [4] for Brouer's method are for the case where ships' speeds are not given in advance and the operated rotations can be with a bi-weekly frequency, which is different from the case of our problem. We therefore have to implement Brouer's method and adapt it for our problem. For both our branch-and-price algorithm and Brouer's method, we set the time limit to be two hours, and for Brouer's method, we further set the limit for the number of iterations to be 10000.

The results are reported in Table 2. Since Brouer's method is only a heuristic method, no upper bounds can be obtained from it to evaluate the optimality gaps. Therefore, we state "n/a" in its column "GAP". From Table 2, it can be seen that *our branch-and-price algorithm always produce better solutions than Plum's and Brouer's methods, and that the obtained improvements are substantial.* Compared with Plum's method, *our branch-and-price algorithm has also produced significantly better optimality gaps, ranging from 2% to 18.5%.* It is worth noting that for some instances, the objective values of the solutions produced by our branch-and-price algorithm are even bigger than the upper bounds reported in [9] for Plum's method. This is because for the model in [9], the maximum number of rotations allowed to be operated is limited, so that their upper bounds

Table 2: Comparison with Plum's and Brouer's methods for solving the benchmark instances

		Plum			Brouer			Branch-and-price		
INST	COL	OBJ	GAP(%)	TIME	OBJ	GAP(%)	TIME	OBJ	GAP(%)	TIME
p12-d22-b-l	385323	427485	30.0	3600	425687	n/a	820	662766	3.6	7200
p12-d22-b-b	932654	408771	39.0	3600	490681	n/a	892	748618	2.0	7200
p12-d22-b-h	955985	636152	3.2	3600	556698	n/a	1516	785773	2.1	7200
p20-d37-w-l	$> 10^9$	1940817	67.9	10800	3902718	n/a	7200	5106296	18.5	7200
p20-d37-w-b	$> 10^9$	3372618	47.3	10800	4278420	n/a	7200	5523499	18.3	7200
p20-d37-w-h	$> 10^9$	3899767	41.0	10800	4456830	n/a	7200	6009782	13.8	7200

are not valid when evaluating our solutions, which are derived without such restrictions.

References

[1] J. F. Álvarez. Joint routing and deployment of a fleet of container vessels. *Maritime Economics & Logistics*, 11(2):186–208, 2009.

[2] J. Andersen, M. Christiansen, T. G. Crainic, and R. Grohaug. Branch and price for service network design with asset management constraints. *Transportation Science*, 41(1):33–49, 2011.

[3] J. Andersen, T. G. Crainic, and M. Christiansen. Service network design with asset management: Formulations and comparative analyses. *Transportation Research Part C: Emerging Technologies*, 17:197–207, 2009.

[4] B. D. Brouer, J. F. Álvarez, C. E. M. Plum, D. Pisinger, and M. M. Sigurd. A base integer programming model and benchmark suite for liner-shipping network design. *Transportation Science*, 48(2):281–312, 2014.

[5] T. G. Crainic, M. Hewitt, M. Toulouse, and D. M. Vu. Service network design with resource constraints. *Transportation Science*, 2014.

[6] Maersk-Line. www.maerskline.com, 2016. Accessed June 7, 2016.

[7] Q. Meng, S. Wang, H. Andersson, and K. Thun. Containership routing and scheduling in liner shipping: overview and future research directions. *Transportation Science*, 48(2):265–280, 2014.

[8] J. Mulder and R. Dekker. Methods for strategic liner shipping network design. *European Journal of Operational Research*, 235(2, SI):367–377, 2014.

[9] C. E. M. Plum, D. Pisinger, and M. M. Sigurd. A service flow model for the liner shipping network design problem. *European Journal of Operational Research*, 235(2):378–386, 2014.

[10] Y. Qu and B. F. Bard. A branch-and-price-and-cut algorithm for heterogeneous pickup and delivery problems with configurable vehicle capacity. *Transportation Science*, 49(2):254–270, 2015.

[11] S. Wang and Q. Meng. Liner ship fleet deployment with container transshipment operations. *Transportation Research Part E: Logistics and Transportation Review*, 48(2):470–484, 2012.

[12] C. Yan and J. Kung. Robust aircraft routing. *Transportation Science*, Article in Advance, 2016.

Planning for charters: a stochastic maritime fleet composition and deployment problem

Xin Wang[1], Kjetil Fagerholt[1], and Stein W. Wallace[2]

[1]Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[2]Norwegian School of Economics (NHH), Bergen, Norway

1 Problem Statement

This study considers a real-life chartering problem faced by Odfjell, a Norwegian public listed company based in Bergen, Norway. As a leading company in the global market for the transportation and storage of bulk liquid chemicals, Odfjell provides services on trading routes all around the world. Each year by the end of October, Odfjell determines the time-charter contracts to enter into for the following year as a supplement to the capacity of the fleet they currently own. On a time charter, a daily hire is paid to the ship owner while the shipping company also bears the sailing costs including fuel and port/canal fees etc. These time charters represent a significant portion of Odfjell's annual expenses and will decide how many ships of each type to charter in, and for how long they are to be hired.

Several aspects of the future market, such as customer demands, can be highly uncertain. For example, some transport contracts only state percentages of the customers' actual production rather than absolute amounts, which make the committed volumes needing transport uncertain. As a result of these uncertainties, the imbalances between supplies and demands for transport capacity in different regions are common in chemical shipping. But this also results in possibilities of picking up optional cargoes from the spot market. Therefore, with the underlying market uncertainties (such as contractual demands and the size of the spot market) affecting the shipping capacity required, the decision making on charters has become rather complicated.

This chartering problem can be seen as a *tactical fleet composition problem* with a focus on capacity adjustment given an existing fleet (Hoff et al., 2010). However, without taking into consideration the operational details to some degree, fleet composition decisions may be based on a too simplified view. Hence, an integration of deployment or routing into the fleet composition decisions is warranted in most cases. In this study, we include fleet deployment decisions to support the capacity evaluation necessary for the making of the charter plan.

The contribution of this study is to present a novel stochastic programming model for the chartering problem, taking into account some of the uncertainties affecting the market. These include stochastic demands, fuel prices, charter rates and freight rates. Even though the model is rather general and applicable to many shipping segments and companies, we demonstrate its use to the case of Odfjell, and focus on the decisions for time charters. We show how the charter plans change as we alter some of the modeling: we vary the level of detail in the modeling of fleet deployment; we use the deterministic version of the original stochastic model; we assume uncorrelated random variables; and we treat speed optimization in different ways. We also show how different chartering plans affect the company's overall performance, in order to provide guidance in helping the company make its chartering decisions.

2 Stochastic Fleet Composition and Deployment Model

The planning period of the charter plan is divided into two periods, **P-1** (January to March) and **P-2** (April to December). The shipping company is quite sure of the demands (contractual and spot) they are facing and is also confident about its prediction on fuel prices, spot rates etc., for **P-1**; but much less so for **P-2**, due to high market volatility. The charter plan then consists of two sub-decisions: the first determines before **P-1** how many and what types of ships to charter in for the next year; and the second makes further adjustments to the chartered-in ships, between **P-1** and **P-2**, by determining whether to increase or decrease the charters for **P-2**. With a fleet of owned and chartered-in ships, the fleet deployment decisions, taking into account speed optimization (see Andersson et al., 2015), allocate shipping capacity to *loops*. Each loop is defined as a round-trip route servicing a number of *trade lanes* that start and end in the same geographic area, where a trade lane represents a transportation arrangement from one geographic area to another which contains one or more *contracts*. The mathematical formulation of the scenario-based two-stage stochastic model is as follows, with the notation shown in Table 1.

Table 1: Notation

Sets	
$\mathcal{V}, \mathcal{K}, \mathcal{C}$	the set of ship types, capacity types and contracts, respectively.
$\mathcal{N}, \mathcal{R}$	the set of trade lanes and loops, respectively.
$\mathcal{E}_v$	the set of speed alternatives for ship type v.
$\mathcal{R}_v \subseteq \mathcal{R}$	the set of loops that can be sailed by ship type v.
$\mathcal{R}_{iv} \subseteq \mathcal{R}$	the set of loops servicing trade lane i that can be sailed by ship type v.
$\mathcal{C}_i \subseteq \mathcal{C}$	the set of contracts serviced by trade lane i.
$\mathcal{V}_i \subseteq \mathcal{V}$	the set of ship types that can sail trade lane i.
$\mathcal{C}_k \subseteq \mathcal{C}$	the set of contracts compatible with capacity type k.
$\mathcal{K}_c \subseteq \mathcal{K}$	the set of capacity types compatible with contract c.
$\mathcal{S}$	the set of scenarios.
Deterministic Parameters	
N_v	no. of ships of type v owned by the shipping company.

M_v^1, M_v^2	no. of available service days for a ship of type v during **P-1** and **P-2**, respectively.
Q_{vk}	volume of capacity type k installed on ship type v.
T_{vre}	total travel time for ship type v to complete a round trip on loop r with speed alternative e, including sailing time and time spent at ports, etc.
F_c^1, F_c^2	frequency requirement of contract c in **P-1** and **P-2**, respectively.
D_c	demand of contract c in **P-1**.
C_{vre}^{RT}	cost for ship type v to complete a round trip on loop r with speed alternative e in **P-1**, including fuel cost, port fees, canal tolls, etc.
C_v^I	daily charter-in rate for a ship of type v on a "long-term" charter (**P-1** plus **P-2**).
$C_v^{\ominus}, C_v^{\oplus}$	(both positive values) adjusting factors for "short-term" charters, representing the additional daily charter-in rate for ship type v if hired only for **P-1**, and only for **P-2**, respectively.
R_v^O	revenue of chartering out a ship of type v per day in **P-1**.
R_{ik}^{SP}	revenue of delivering one unit of spot cargo with capacity type k on leg i in **P-1**.
D_{ik}^{SP}	no. of units of spot cargo available on trade lane i that are compatible with capacity type k in **P-1**.
Stochastic Parameters	
p^s	the probability of scenario s taking place in **P-2**.
D_{cs}	demand of contract c for scenario s in **P-2**.
C_{vres}^{RT}	cost for ship type v to complete a round trip on loop r with speed alternative e for scenario s in **P-2**.
C_{vs}^I	cost of chartering in a ship of type v per day for scenario s in **P-2** ("on the spot" extra time charters).
R_{vs}^O	revenue of chartering out a ship of type v per day for scenario s in **P-2**.
R_{iks}^{SP}	revenue of delivering one unit of spot cargo with capacity type k on trade lane i for scenario s in **P-2**.
D_{iks}^{SP}	no. of units of spot cargo available on trade lane i that are compatible with capacity type k for scenario s in **P-2**.
Decision Variables	
w_v	(charter plan) no. of ships of type v chartered in at the start of **P-1**.
$w_v^{\ominus}, w_v^{\oplus}$	(charter plan) no. of ships of type v to reduce or add (based on w_v), respectively, at the end of **P-1**.
x_{vre}, x_{vres}	no. of round trips sailed by a ship of type v on loop r with speed alternative e in **P-1**, and for scenario s in **P-2**.
y_{vs}	no. of days of extra charter-in for ship type v in scenario s in **P-2**.
z_v, z_{vs}	no. of days of chartering out ship type v in **P-1**, and for scenario s in **P-2**.
q_{ivkc}, q_{ivkcs}	volume of contract c carried by capacity type k installed on ship type v on trade lane i in **P-1**, and for scenario s in **P-2**.
$q_{ivk}^{SP}, q_{ivks}^{SP}$	volume of spot cargo carried by capacity type k installed on ship type v on trade lane i in **P-1**, and for scenario s in **P-2**.

$$\min \quad \sum_{v \in \mathcal{V}} \Big(C_v^I M_v^1 w_v + C_v^I M_v^2 (w_v - w_v^{\ominus} + w_v^{\oplus}) + C_v^{\ominus} M_v^1 w_v^{\ominus} + C_v^{\oplus} M_v^2 w_v^{\oplus} \Big) \tag{1.a}$$

$$+ \sum_{v \in \mathcal{V}} \sum_{r \in \mathcal{R}_v} C_{vre}^{RT} x_{vre} - \sum_{v \in \mathcal{V}} R_v^O z_v - \sum_{i \in \mathcal{N}} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}} R_{ik}^{SP} q_{ivk}^{SP} \tag{1.b}$$

$$+ \sum_{s \in \mathcal{S}} p^s \Big(\sum_{v \in \mathcal{V}} \sum_{r \in \mathcal{R}_v} C_{vres}^{RT} x_{vres} + \sum_{v \in \mathcal{V}} C_{vs}^I y_{vs} - \sum_{v \in \mathcal{V}} R_{vs}^O z_{vs} - \sum_{i \in \mathcal{N}} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}} R_{iks}^{SP} q_{ivks}^{SP} \Big) \tag{1.c}$$

s.t.

$$\sum_{r\in\mathcal{R}_v}\sum_{e\in\mathcal{E}_v} T_{vre}x_{vre} + z_v = M_v^1(N_v + w_v) \qquad v\in\mathcal{V} \tag{1}$$

$$\sum_{v\in\mathcal{V}_i}\sum_{r\in\mathcal{R}_{iv}}\sum_{e\in\mathcal{E}_v} x_{vre} \geq F_c^1 \qquad i\in\mathcal{N}, c\in\mathcal{C}_i \tag{2}$$

$$\sum_{v\in\mathcal{V}_i}\sum_{k\in\mathcal{K}_c} q_{ivkc} = D_c \qquad i\in\mathcal{N}, c\in\mathcal{C}_i \tag{3}$$

$$\sum_{r\in\mathcal{R}_{iv}}\sum_{e\in\mathcal{E}_v} Q_{vk}x_{vre} \geq \sum_{c\in\mathcal{C}_i\cap\mathcal{C}_k} q_{ivkc} + q_{ivk}^{SP} \qquad i\in\mathcal{N}, v\in\mathcal{V}_i, k\in\mathcal{K} \tag{4}$$

$$\sum_{v\in\mathcal{V}_i} q_{ivk}^{SP} \leq D_{ik}^{SP} \qquad i\in\mathcal{N}, k\in\mathcal{K} \tag{5}$$

and (2^{nd}-stage constraints)

$$\sum_{r\in\mathcal{R}_v}\sum_{e\in\mathcal{E}_v} T_{vre}x_{vres} + z_{vs} = M_v^2(N_v + w_v + w_v^{\oplus} - w_v^{\ominus}) + y_{vs} \quad v\in\mathcal{V}, s\in\mathcal{S} \tag{6}$$

$$\sum_{v\in\mathcal{V}_i}\sum_{r\in\mathcal{R}_{iv}}\sum_{e\in\mathcal{E}_v} x_{vres} \geq F_c^2 \qquad i\in\mathcal{N}, c\in\mathcal{C}_i, s\in\mathcal{S} \tag{7}$$

$$\sum_{v\in\mathcal{V}_i}\sum_{k\in\mathcal{K}_c} q_{ivkcs} = D_{cs} \qquad i\in\mathcal{N}, c\in\mathcal{C}_i, s\in\mathcal{S} \tag{8}$$

$$\sum_{r\in\mathcal{R}_{iv}}\sum_{e\in\mathcal{E}_v} Q_{vk}x_{vres} \geq \sum_{c\in\mathcal{C}_i\cap\mathcal{C}_k} q_{ivkcs} + q_{ivks}^{SP} \qquad i\in\mathcal{N}, v\in\mathcal{V}_i, k\in\mathcal{K}, s\in\mathcal{S} \tag{9}$$

$$\sum_{v\in\mathcal{V}_i} q_{ivks}^{SP} \leq D_{iks}^{SP} \qquad i\in\mathcal{N}, k\in\mathcal{K}, s\in\mathcal{S} \tag{10}$$

variable domains

$$w_v, w_v^{\oplus}, w_v^{\ominus} \geq 0 \qquad v\in\mathcal{V} \tag{11}$$

$$w_v \geq w_v^{\ominus} \qquad v\in\mathcal{V} \tag{12}$$

$$x_{vr}, x_{vrs}, y_{vs}, z_v, z_{vs} \geq 0 \qquad v\in\mathcal{V}, r\in\mathcal{R}_v, s\in\mathcal{S} \tag{13}$$

$$q_{ivkc}, q_{ivkcs} \geq 0 \qquad i\in\mathcal{N}, v\in\mathcal{V}_i, k\in\mathcal{K}, c\in\mathcal{C}_i\cap\mathcal{C}_k, s\in\mathcal{S} \tag{14}$$

$$q_{ivk}^{SP}, q_{ivks}^{SP} \geq 0 \qquad i\in\mathcal{N}, v\in\mathcal{V}_i, k\in\mathcal{K}, s\in\mathcal{S} \tag{15}$$

The objection function minimizes the sum of the chartering costs (1.a), the operating costs in **P-1** (1.b) and the expected operating costs in **P-2** (1.c). Constraints (1) state that all transport availability of the fleet is used either through the carrier's own operations or chartered out. Constraints (2) ensure the satisfaction of the frequency requirement of every contract and Constraints (3) the demand requirement. Constraints (4) ensure that the total volume of capacity type k installed on ship type v sailing on trade lane i is respected, and may be used to carry either contractual or spot cargoes. Constraints (5) restrict the amount of spot cargo carried by the shipping company within the size of the spot market for the respective capacity type. Constraints (6) - (10) are the stochastic **P-2** versions of constraints (1) - (5) for the second stage.

3 Computational Study

We present a case study based on realistic data from Odfjell, with a network of 9 geographic areas and 22 trade lanes. We first study the effects of altering the level of detail in the modeling of fleet deployment by changing the maximum number of trade lanes making up a loop. Our results show that increasing the highest loop cardinality from two to three leads to a 6.8% reduction in total costs, but when increasing such value from three to four or higher the improvements are less significant (less than 1%) while the run times grow dramatically (solved with CPLEX). It is further shown that those loops with three trade lanes and low ballast sailing ratios are potentially of great value to the shipping company.

We then evaluate the charter plans produced using (1) the deterministic model where all uncertain parameters are replaced by their means (as most companies do in the industry), (2) the deterministic model with higher demand expectation and (3) the stochastic model but without considering correlations among the stochastic demands, freight rates and charter rates, etc. Our results show that using the deterministic model with mean values gives a loss of almost 13% in total costs compared to the optimal stochastic solution, and such loss can be brought down to around 5% if higher deterministic demand values are used as forecast. Also, disregarding correlation information leads to a loss of 4.9% even if the stochastic model is used. Therefore, the shipping company should, where possible, use the stochastic model and take both individual distributions and correlation information into account; and plan with higher demand expectation than the means if the company has to use the deterministic model. However, due to the incompatibilities between trade lanes and ship types, and between contracts and capacity (tank) types, deterministic models often struggle with providing the "correct" combination (mix) of the different types of ships to charter in.

In addition, we investigate how much we can gain from taking speed optimization into account when making the charter decisions, since in practice this type of tactical plan is usually made assuming one (design) speed for each ship type. Our results show that integrating speed optimization in the stochastic fleet composition and deployment model is beneficial and leads to an improvement of almost 10% percent in total costs.

References

Andersson, H., Fagerholt, K., and Hobbesland, K. (2015). Integrated maritime fleet deployment and speed optimization: Case study from RoRo shipping. *Computers & Operations Research*, 55:233–240.

Hoff, A., Andersson, H., Christiansen, M., Hasle, G., and Løkketangen, A. (2010). Industrial aspects and literature survey: Fleet composition and routing. *Computers & Operations Research*, 37(12):2041–2061.

Multimodal Transportation Services

SA1: Rail Transport

Saturday 9:00 – 11:00 AM

Session Chair: Dario Pacciarelli

9:00 The Load Planning Problem for Double-Stack Trains at Intermodal Terminals

Serena Mantovani, Gianluca Morganti, Nitish Umang, Teodor Gabriel Crainic, Emma Frejinger, Eric Larsen*

CIRRELT and Université de Montréal

9:30 Tactical Block and Car Planning for Intermodal Trains

[1]Gianluca Morganti, [2]Teodor Gabriel Crainic, [1]Emma Frejinger, [3]Nicoletta Ricciardi*

[1]CIRRELT and Université de Montréal, [2]ESG, UQAM and CIRRELT, [3]Sapienza University of Rome

10:00 Optimization of Handouts for Rolling Stock Rotations Visualization

Boris Grimm, Ralf Borndoerfer, Thomas Schlechte, Markus Reuther*

Zuse Institute Berlin

10:30 Real-Time Near-Optimal Train Scheduling and Routing in Complex Railway Networks

[1]Marcella Samà, [1]Andrea D'Ariano, [1]Dario Pacciarelli, [2]Francesco Corman*

[1]Roma Tre University, [2]Delft University of Technology

The load planning problem for double-stack trains at intermodal terminals

Serena Mantovani * † Gianluca Morganti †
Nitish Umang † Teodor Gabriel Crainic † ‡
Emma Frejinger † Eric Larsen †

1 Introduction

In this work, we present a general methodology that addresses the load planning problem for double-stack intermodal trains. It arises at intermodal rail terminals that are major components of any intermodal transportation system providing space and equipment for, e.g., classifying, storing and unloading/loading containers. The problem consists in assigning containers to outbound railcars where each railcar has one or several platforms (each platform has in turn one or two slots: bottom and, in case of double-stack, top slot). This is challenging because there are many different railcar and container types which give rise to complicated loading rules that depend on the characteristics of both. The proposed methodology can be used to provide decision support to terminal managers.

Most of the studies in the literature focus on the single-stack load planning problem. In this case, the loading problem is rather simple and the aim is to minimize handling costs in the yard (e.g., Corry and Kozan, 2008) or train set-up costs (Bruns and Knust, 2012). To the best of our knowledge, Lai et al. (2008) is the only study dealing with the double-stack load planning problem. A number of important aspects are, however, ignored in this work. First, they address the matching among containers and railcar types but assume that it is independent over platforms, which is not true in our case. Second, center-of-gravity restrictions are not considered. We contribute to the literature by providing a general

*Corresponding author, email: serena.mantovani@cirrelt.ca
†Department of Computer Science and Operational Research, Université de Montréal and CIRRELT, Canada
‡Départément management et technologie, École des Sciences de la Gestion, Université du Québec a Montréal, Canada

methodology that can address the load planning problem accounting for all crucial loading constraints. We present numerical results for the North American market.

Given a set of containers stored in the terminal and a departing train, the problem we consider consists in selecting the optimal set of containers to load and the optimal way of loading them, using the maximum of the available capacity. We address this problem for double-stack trains, where the load planning problem is challenging because of a number of loading rules that depend on container and railcar characteristics, and on the way they match together. The size, the location of the load-bearing along the length of the container, the type (e.g., tanker and dangerous-goods containers have restrictions with respect to the position in the stack they may occupy) of containers determine how containers can be stacked.

Intermodal trains are composed of sequences of railcars, designed to carry single- or double-stacked containers. Railcars differ on attributes such as the number and the length of platforms and the weight holding limit. Loading rules depend on these characteristics. For the North American market, rules for matching containers and railcar types are presented in the AAR guide (Association American Railroads, 2014). It provides information on the container sizes that can be loaded in the bottom and top slot of each platform. We refer to them as *containers-to-cars matching rules*. We note that the guide reports the loading capabilities but it does not show all the possible ways of matching. Moreover, the loading patterns for certain platforms may depend on the loading of the others, thus, it is not possible to decompose the loading rules by platforms.

2 Methodology

We model the matching among types of containers and railcars through loading patterns. This is similar to Corry and Kozan (2008) and Lai et al. (2008), but we account for dependencies between the loading of the platforms of the same railcar. This leads to sets of loading patterns with large cardinality. We also model weight holding restrictions. First, the total weight of the containers loaded on a given platform cannot exceed its weight holding capacity. Second, the vertical center of gravity cannot exceed a given threshold which imposes a maximum weight limit on containers loaded in top slots.

We present an integer linear programming (ILP) formulation, where we have two types of *decision variables*. First, binary variables assigning containers to slots and second, binary variables assigning loading patterns to railcars. The objective of the model seeks to minimize the total cost of the containers left on the ground and of the used railcars that are assigned at least one container. This generalized cost leads to the maximization of the slot utilization on the selected railcars. The ILP formulation has a number of different types of constraints: assignment and loading pattern constraints, weight capacity and center-of-gravity constraints as well as technical loading constraints and container-cargo specific constraints including stacking restrictions.

3 Results

We present two numerical studies. The first one has the purpose to assess the importance of modeling containers-to-cars matching and center-of-gravity restrictions. The second one, to analyze the computational time for large size instances with different characteristics. Both studies use simulated data but with real railcars (sampled from the types in the AAR Guide) and containers with realistic characteristics.

In the first study, we fix the train length to 5,000 ft and we vary the composition of the railcars (four different types: 40 ft one/five platforms and 53 ft one/five platforms). We build container sets by varying the proportion of 40 ft and 53 ft containers in the instances. Moreover, we create three sets of containers using different weight assumptions: two deterministic settings that should not have center-of-gravity issues (all containers have equal weight or 50% are light and 50% are heavy) and 20 instances with random weight (drawn from empirical distribution).

250 CONTAINERS AVAILABLE	25 FIVE 40' PLATFORM RAILCARS AVAILABLE		
INSTANCE DESCRIPTION	LOADED CONTAINERS	USED RAILCARS	SOLUTION TIME [sec]
S1: 200 40ft containers and 50 53ft containers			
1) Containers same weights	250	25	126.59
2) Containers half low and half high weights	250	25	132.83
3) Containers random weights	233	23.75	935.88
S2: 100 40ft containers and 150 53ft containers			
1) Containers same weights	175	25	113.67
2) Containers half low and half high weights	175	25	125.89
3) Containers random weights	175	25	329.88
200 CONTAINERS AVAILABLE	**20 FIVE 53' PLATFORM RAILCARS AVAILABLE**		
INSTANCE DESCRIPTION	LOADED CONTAINERS	USED RAILCARS	SOLUTION TIME [sec]
S3: 200 40ft containers			
1) Containers same weights	200	20	514.81
2) Containers low and high weights	200	20	539.24
3) Containers random weights	200	20	733.83
S4: 0 40ft containers and 200 53ft containers			
1) Containers same weights	200	20	429.88
2) Containers low and high weights	200	20	513.44
3) Containers random weights	200	20	859.54

Table 1: Importance of the matching problem and the center of gravity, five-platform railcars: number of loaded containers, used railcars and solution time.

Because of limited space, we present only the results for five platform railcars in Table 1. The first two columns show the number of loaded containers and the number of used railcars in the optimal solution. The third column shows CPLEX solution time. We start the analysis of the results firstly focusing on the container-to-car matching effect. In case of 40 ft platform railcars (settings S1 and S2), the maximum number of containers that can be loaded on 250 slots is 250, but 53 ft containers can only be loaded in top slots because of the platforms' length. As long as the weights are "favorable" (respects center of gravity) all available top slots can be used to load 53 ft containers. The results show that this is not the case for random weights. Indeed, there is a drop in the average number of loaded containers, even when the sizes match well.

The 53 ft platform railcars (settings S3 and S4) are more flexible since they can take 53 ft containers in any slot, but because of the railcar length, the maximum capacity is less than for 40 ft railcars. Under all the weight conditions, it is possible to load all the containers by simply placing the heaviest containers in bottom slots.

In a second numerical study, we draw railcars from the available types in the North American fleet as well as containers of different sizes and weights. We use three different train lengths: 2,000 ft (representative example for many countries across the world), 6,000 ft (long train) and 10,000 ft (extremely long train). For each of the train lengths we generate 20 sequences of railcars by sampling railcar types. For each train length, and for each railcar sequence, we sample 10 sets of containers, where the set size is 1.5 times the number of slots in the railcar sequence. Given the excess demand, we expect to always achieve a near 100 % slot utilization. The results reported in 2 show that all instances can be solved to optimality in reasonable computational time. It takes considerably longer time to solve instances with all container sizes (20, 40, 45, 48 and 53) compared to instances with only 20 and 40 ft containers. This is due to the cardinality of the sets of loading patterns.

	Containers without technical loading restrictions		**Containers with technical loading restrictions**	
Train length (ft)	**20ft and 40ft**	**All sizes**	**20ft and 40ft**	**All sizes**
2,000	7.11	13.10	7.95	14.36
6,000	184.59	450.96	639.37	661.04
10,000	967.42	4,010.52	1,963.78	8,266.35

Table 2: Average computational time in seconds

4 Conclusion

We presented a general methodology that addresses the load planning problem for double-stack intermodal trains. We model containers-to-cars matching rules, center-of-gravity constraints, stacking rules and technical loading restrictions associated with specific container types and/or goods. We formulated the problem as an integer linear programming (ILP) model. Results showed that ignoring containers-to-cars matching and center-of-gravity restrictions may lead to an overestimation of the train capacity and to infeasible load plans. The results also showed that we can solve realistic instances to optimality in reasonable time.

The model and results have not been presented in a conference before. A substantially different version of the model and other results were presented at the ODYSSEUS conference in 2015. The article was submitted to EJOR on December 18, 2016.

Acknowledgements

This research was funded by Mitacs and a Collaborative Research and Development Grants, Natural Sciences and Engineering Research Council of Canada (CRD-477938-14). While working on this project, the fifth author was holding a Chair on Intermodal Transportation at Université de Montréal, the third was postdoctoral fellow with the Chair, and the fourth was Adjunct Professor with the Department of Computer Science and Operations Research, Université de Montréal. The authors present their most sincere thanks Mr. Serge Bisaillon, analyst at CIRRELT, for his invaluable support. We are also thankful to Jean-François Cordeau for helpful discussions. Finally, we gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants.

References

Association American Railroads (2014). Loading capabilities guide. https://www.aar.org/StatisticsAndPublications/Publications?title=LoadingCapabilitiesGuide.

Bruns, F. and Knust, S. (2012). Optimized load planning of trains in intermodal transportation. *OR Spectrum*, 34(3):511–533.

Corry, P. and Kozan, E. (2008). Optimised loading patterns for intermodal trains. *OR Spectrum*, 30(4):721–750.

Lai, Y.-C., Barkan, C. P., and Önal, H. (2008). Optimizing the aerodynamic efficiency of intermodal freight trains. *Transportation Research Part E*, 44(5):820–834.

Tactical block and car planning for intermodal trains

Gianluca Morganti [*] [†] Teodor Gabriel Crainic [†] [‡]
Emma Frejinger [†]

1 Introduction

Railroads are at the heart of freight transportation systems, moving a broad variety of commodities long distances in a cost-effective and environmentally sustainable manner. They are, in particular, a key element of the world-wide intermodal transportation network, which displays a steady traffic growth, as illustrated by the the 5.3% annual increase (average, since 1990) at North American ports (IAPH, 2015).

Efficient and profitable railroad activities require adequate planning of operations and resources. We focus on the *block and car tactical planning problem* that arises in intermodal rail transportation, using the North American market as a case study. There are several studies in the literature focusing on the train blocking problem but the existing models do not directly apply to intermodal trains. We thus propose a new model that accounts for demand expressed in terms of numbers of containers of various types and explicitly treats the assignment of these containers to railcars of different types. The model then assigns railcars to blocks and blocks to trains. In the following, we provide some background on intermodal rail transport to motivate our work, as well as a brief literature review. We present a problem description and the model in Section 2, and sum up the presentation plan in Section 3.

While intermodal traffic makes use of the same tracks and in certain cases the same equipment (e.g.,locomotives) as trains moving other types of cargo, it entails important differences. The most obvious one is that intermodal demand is expressed in numbers

[*]Corresponding author, email: gianluca.morganti@cirrelt.ca
[†]Department of Computer Science and Operational Research, Université de Montréal and CIRRELT, Canada
[‡]Départément management et technologie, École des Sciences de la Gestion, Université du Québec a Montréal, Canada

of containers, while it is tonnage translated into numbers of railcars of particular types in the other case. About 90% of the containers used worldwide are either 20 or 40 feet, while longer units (45, 48 and 53 feet) are also used in the North American market (Wikipedia, 2016). This give rise to a demand structure where shares of the different types of containers to transport vary over origin-destination (OD) pairs. In terms of equipment, intermodal railcars are different from other railcars. Cars with one to five platforms are used in North America and they can often be double-stacked. The railcars have different loading capabilities depending on the number of platforms, their length and weight holding capacity. A railcar with five 40-foot platforms can, for example, transport a maximum of ten 40-foot containers or five 40-foot and three 53-foot containers. The multitude of railcar and container types results in a large number of ways to load containers that must respect different loading rules (Mantovani et al., 2016). Representing in a computationally efficient way the assignment of containers to railcars within a tactical blocking model is a particular challenge of the problem we address. A second challenge is to account for the utilization of available railcars, which are generally in short supply.

A block is a group of railcars, with possibly different origins and destinations, that move as a single unit between a pair of yards, without cars being handled individually (e.g., sorted) at intermediate yards. Blocking thus aims to take advantage of economies of scale and reduce the cost of handling railcars at yards. A block is moved by a sequence of trains, while a railcar can be moved by one or a sequence of blocks between its origin and destination yards. The classical *train blocking problem* then consists in selecting the blocks to build and assigning railcars with given OD pairs to blocks. Several studies in the literature focus on the train blocking problem (e.g., Bodin et al., 1980; Newton et al., 1998; Barnhart et al., 2000), while Zhu et al. (2014) integrates blocking within a comprehensive scheduled service network design model, but none of these studies accounts for the container-to-railcar assignment or the management of the railcar fleets (see Crainic et al., 2014, for a general model of the latter issue).

As detailed next, we propose a model that considers several types of containers and railcars, integrates the contaier-to-railcar assignment, and accounts for the utilization of the railcar fleets. The train schedule is given in the problem setting we consider.

2 Problem Description and Modeling

We focus on the problem of blocking intermodal traffic. In this case, the demand (or shipments) corresponds to containers, which are characterized by their type, origin-destination pair, arrival time at origin and required time at the destination. The railroad uses a fleet of intermodal railcars to move the demand. The fleet can be composed of various types of railcars designed to move containers with different characteristics. There is a large variety of railcars on the North American market (Association American Railroads, 2014). Each railcar has a number of platforms and slots on the platforms. Double-stack platforms have two slots while single-stack platforms have one. The matching of containers to slots given the railcar type is an important issue since all slots may not be usable when the container type matches the railcar type poorly. For example, 40-foot platforms cannot take 53-foot containers in the bottom slot. The container types should therefore be defined based on the attributes that influence the loading capabilities of the railcars, e.g., size, cargo type (dry, refrigerated) or weight (heavy or light). We are therefore faced with three consolidation processes, assignment of containers to railcars, of railcars to blocks and of blocks to trains, where the differentiation of railcar and container types is a paramount consideration.

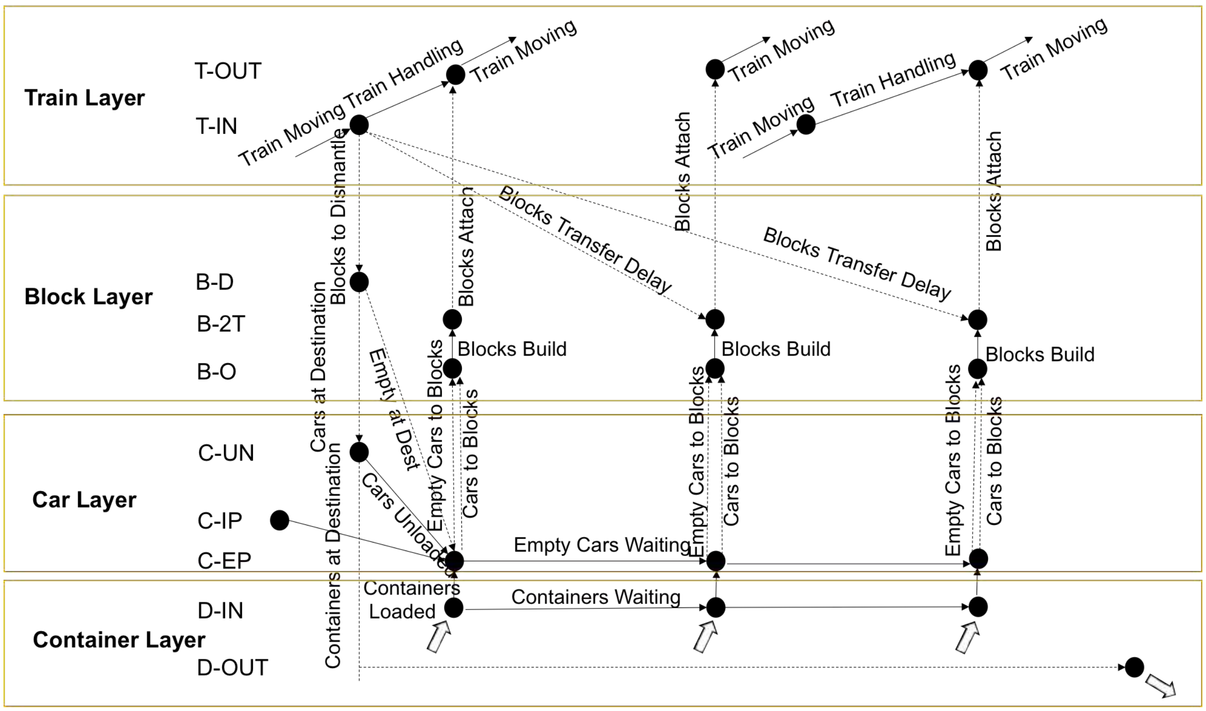


Figure 1: Time-Space Network Representation

We propose a model that is based on a cyclic four-layer space-time network representation, illustrated in Figure 1. This is a tactical problem and the schedule is defined over given schedule length (e.g., a week) and is assumed to be repeated over a planning horizon. Since intermodal traffic shares the network with trains moving other types of cargo, we take the train schedule as given. Different from most studies in the literature (e.g., Zhu et al., 2014), this allows us to use a continuous-time network representation. Moreover, we do not have any decision variables related to the *train* (top) *layer*. For each scheduled intermodal train, the arrival and departure times at each terminal are known and yield the *T-IN* and (*T-OUT* nodes, respectively. We model two train activities with different types of arcs: *waiting* arcs that represent the time a train spends in a terminal (T-IN to T-OUT) and *moving* arcs between terminals (T-OUT to T-IN). A path in the train layer corresponds to a train service. Features such as power and maximum length of the train are assumed to be known and give the capacity of the moving arcs.

The *block layer* is illustrated below the train layer. We generate the nodes and arcs in the block layer based on the nodes in the train layer, and a *block* with a particular (time-dependent) OD pair is represented by a *path* in the block layer. We model three activities: *building* new blocks, *transferring* blocks from one train to another, and *dismantling* blocks at their destinations. These activities are represented by three sets of nodes (B-IN, B-2T and B-OUT) and arcs. The *Blocks Build* arcs (B-O to B-2T) are the only in-layer arcs, but there are three types of inter-layer arcs between the block and train layers: *Blocks Attach* (attaching a block to a train), *Blocks to Dismantle* (blocks that have arrived at their destination are disconnected from the train) and *Blocks to Transfer Delay* (blocks are disconnected from a train and wait in the terminal for their next train).

The *car layer* (situated between the container and block layers) is used to model the activities of assigning containers to railcars and railcars to blocks. The fleet of available empty railcars is represented by a source node at the beginning of the time horizon at each terminal (C-IP). There are also nodes representing the start time of container-to-railcar assignment (C-EP) and the start time of the container unloading activities (C-UN). Several important types of constraints are associated with the C-EP nodes because this is where the matching of demand (different types of containers that should be loaded, represented by an inter-layer arc) and empty railcars take place. The empty railcars can come from the pool (C-IP), empty railcars waiting from a previous time period, from railcars that have been unloaded or as railcars that arrived empty (repositioning represented by an inter-layer arc). The demand is represented in the *container laye*r where the continuous arrival of containers is accumulated over time intervals (e.g.. daily). A sink node called D-OUT accumulates all the containers arrived at the destination and is connected to C-UN by inter-layer arcs.

We propose a mixed integer linear programming (MILP) formulation with four groups of decision variables: (i) *Block selection*, a binary variable equal to 1 if a block is selected or 0 otherwise; (ii) *Container flow distribution* representing the number of different types of containers on a given arc in the container layer; (iii) *Loaded railcar flow distribution* representing the number of loaded railcars of a certain type transporting a number of containers of a given type (demand) on a given arc; (iv) *Empty railcar flow distribution* representing the number of empty railcars of a given type on a given arc. The objective minimizes the total cost of the system over the planning horizon. It encompasses the cost of selecting, operating and transferring blocks as well as the penalty for late arrival of demand and the cost of utilizing resources. There are flow conservation constraints in the four layers, linking constraints, assignment constraints and yard capacity constraints. The latter are defined as bundle constraints regarding the maximum length of blocks that can be built and dismantled during a given time interval at a given terminal.

3 Conclusion

We address the blocking problem of intermodal trains in the North American market. The problem is complex dues to the many different railcar and container types, which require to model the container-to-railcar assignment and the utilization of multiple types of resources. This makes the problem different and more complicated than the train blocking problems considered in the literature. We propose a MILP formulation based on a continuous-time, multi-layer network. We take advantage of the problem characteristics, e.g., the known train schedule, to build a solution method that considerably reduces the number of feasible blocks and may thus use a commercial solver. We will present and discuss the model, solution method and numerical results for realistic instances.

References

Association American Railroads (2014). Loading capabilities guide. https://www.aar.org/StatisticsAndPublications/Publications?title=LoadingCapabilitiesGuide.

Barnhart, C., Jin, H., and Vance, P. H. (2000). Railroad blocking: A network design application. *Operations Research*, 48(4):603–614.

Bodin, L. D., Golden, B. L., Schuster, A. D., and Romig, W. (1980). A model for the blocking of trains. *Transportation Research Part B: Methodological*, 14(1):115–120.

Crainic, T. G., Hewitt, M., Toulouse, M., and Vu, D. M. (2014). Service network design with resource constraints. *Transportation Science*, 50(4):1380–1393.

IAPH (2015). World port data 2015. International Assocaciation of Ports and Harbors, http://www.iaphworldports.org/statistics. [Online; accessed 20-September-2016].

Mantovani, S., Morganti, G., Umang, N., Crainic, T. G., Frejinger, E., and Larsen, E. (2016). The load planning problem for double-stack intermodal trains. Technical Report 2016-12-, CIRRELT.

Newton, H. N., Barnhart, C., and Vance, P. H. (1998). Constructing railroad blocking plans to minimize handling costs. *Transportation Science*, 32(4):330–345.

Wikipedia (2016). Intermodal containers. https://en.wikipedia.org/w/index.php?title=Intermodal_container&oldid=737261445. [Online; accessed 02-September-2016].

Zhu, E., Crainic, T. G., and Gendreau, M. (2014). Scheduled service network design for freight rail transportation. *Operations Research*, 62(2):383–400.

Optimization of Handouts for Rolling Stock Rotations Visualization

Ralf Borndörfer, Boris Grimm, Markus Reuther, Thomas Schlechte[1]

1 **Zuse Institute Berlin**
Takustrasse 7, 14195 Berlin, Germany
`{borndoerfer, grimm, reuther, schlechte}@zib.de`

1 Introduction

This paper deals with the problem of producing optimized visualizations of railway rolling stock rotations. The problem arises during the task of assigning railway rolling stock vehicles to timetabled passenger trips of a cyclic timetable at our industrial partner DB Fernverkehr AG (DBF). DBF is one of the largest passenger railway operators in Europe. So what are *rotations* that should be visualized? Imagine a timetable that is defined for one week, i.e., there are planned passenger trips for each day from Monday to Sunday. We call this timetable the *standard week*. The standard week is usually repeated for several weeks or month. To operate the trips during that period the trips have to be assigned to a physical vehicle that operates it. This defines an optimization problem and its outcome is a cyclic sequence of trips for each vehicle. This cyclic sequence is called *rotation*. Remark, the rotation could pass the standard week $\mathfrak{v} \in \mathbb{Z}_+, \mathfrak{v} > 1$ times without operating a trip twice. In that case n physical vehicles are required to operate the rotation. In reality, it turns out that it is just not enough to compute a solution. We also have to think about how to make it easy to manage. In simple words, we ask how to print a rolling stock rotation on a physical paper? The reason for that it that the rotations could be easily communicated and checked for desired properties with that. There is a fixed format how to visualize rotations at DBF in order to obtain a clear picture of repeated trips. But, there are many different possibilities to visualize the same rotations. This defines the optimization problem which is topic of this paper. We introduce the handout visualization problem for rolling stock rotations and developed an optimization approach, i.e., an IP formulation. Moreover, we design a assignment-based construction heuristic, which turns out to be very fast and powerful compared to lower bounds of the IP formulation. A good visualization of a rolling stock rotation was one of the keys to bring optimized rolling stock rotations into operation because it directly increases user acceptance.

2 Rolling Stock Rotation Handouts

The methodology of rolling stock rotation visualizations in terms of the standard week is standardized across many railway companies. For example, NS Reizigers, the Österreichische Bundesbahn, Trenitalia, and, Deutsche Bahn are using the visualization concept that we call *handout concept*.

2.1 Handout Segments

By construction, each rolling stock rotation runs an integral number of times through the set of operational days $\mathbb{D}$ until it reaches again its first trip. We denote this number by $\mathfrak{v} \in \mathbb{Z}_+$ for the rolling stock rotation that we want to visualize. The number of rolling stock

$\Omega(s)$	$\mathbb{D}(s)$	$s \in \mathcal{S}$	$\Omega(\text{next}(s))$
1	Mon	374 → 1061	2
1	Tue	373 → 376	2
1	Wed	374 → 1061	2
1	Thu	373 → 376	2
1	Fri	374 → 1061	2
1	Sat	373 → 376	2
1	Sun	374 → 1061	2
2	Mon	277 → 994	1
2	Tue	374 → 1061	1
2	Wed	373 → 376	1
2	Thu	374 → 1061	1
2	Fri	373 → 376	1
2	Sat	374 → 1061	1
2	Sun	373 → 376	1

$\Omega(s)$	$\mathbb{D}(s)$	$s \in \mathcal{S}$	$\Omega(\text{next}(s))$
1	Mon	374 → 1061	1
1	Tue	374 → 1061	1
1	Wed	374 → 1061	1
1	Thu	374 → 1061	1
1	Fri	374 → 1061	1
1	Sat	374 → 1061	1
1	Sun	374 → 1061	2
2	Mon	277 → 994	2
2	Tue	373 → 376	2
2	Wed	373 → 376	2
2	Thu	373 → 376	2
2	Fri	373 → 376	2
2	Sat	373 → 376	2
2	Sun	373 → 376	1

Figure 1 Two different handouts for the same rolling stock rotation.

vehicles needed to operate the rotation is also equal to $\mathfrak{v}$, i.e., $\mathfrak{v}$ rolling stock vehicles run through the rotation one by one.

For the visualization, we imagine a rolling stock rotation as a cycle that runs through timetabled trips of a standard week, i.e., $|\mathbb{D}| = 7$. In order to find a convenient visualization, the rotation is split into $|\mathbb{D}| \cdot \mathfrak{v}$ *segments*. A segment represents the operation of the rotation on a single day of operation. A segment may be empty, i.e., does not contain any timetabled trip at all. Note that the assumption $|\mathbb{D}| = 7$ is motivated by the fact that we consider the standard week. Nevertheless, all considerations in this paper also apply to other non-trivial planning horizons. Let S be the set of segments of the rolling stock rotation to be visualized and let $\mathbb{D}(s) \in \{\text{Mon}, \ldots, \text{Sun}\} =: \mathbb{D}$ denote the day of operation of the segment $s \in S$ and the set of days of operation, respectively. In addition, we denote by $[\mathfrak{v}] := \{k \in \mathbb{N} \,|\, k \leq \mathfrak{v}\}$ the set of the first $\mathfrak{v}$ natural numbers. A *handout* is a function $\Omega : S \mapsto [\mathfrak{v}]$ such that

$$\mathbb{D}(s) = \mathbb{D}(t) \quad \Rightarrow \quad \Omega(s) \neq \Omega(t) \quad \forall s, t \in S.$$

That is, Ω assigns different values of $[\mathfrak{v}]$ to each pair of segments that are both associated with the same day of operation. By definition, always exactly $\mathfrak{v}$ segments of a rolling stock rotation are associated with the same day of operation. Thus, if a handout Ω is at hand each segment $s \in S$ can be precisely identified by $\Omega(s)$ and $\mathbb{D}(s)$. This is an evident motivation for the concept of handouts. For most of the rolling stock rotations in industry $\mathfrak{v}$ is much greater than one. Indeed, rolling stock rotations with $\mathfrak{v} > 40$ are not an exception at DBF. For those rotations a handout obviously provides a significant gain: Segments can now be precisely distinguished during all further planning steps. The major objective of a handout Ω is to create the standardized visualization for rolling stock rotations that we mentioned above. Indeed,Ω completely defines this visualization. The visualization appears by printing all segments one below the other such that they are lexicographically ordered according to Ω and the day of operation. Figure 1 provides two different visualizations (i.e., the two tables) derived from two different handout functions for the rolling stock rotation. In both tables the first three columns state the value $\Omega(s)$ for a segment $s \in S$ associated with $\mathbb{D}(s)$. The underlying rolling stock rotation is *not* modified by those visualizations. Therefore, the successor relations of the segments that are defined by the rolling stock rotation remain. Segment $\text{next}(s) \in S$ denotes the direct successor of $s \in S$ in the rotation. This order of precedence is stated in the last columns of the two tables. Thus, the rolling stock rotation that is visualized by a handout Ω can be directly followed in the tables. For example, for the direct successor $t \in S$ of the segment $s \in S$ (i.e., $\text{next}(s) = t$) with $\Omega(s) = 2$ and $\mathbb{D}(s) = \text{Sun}$ on the left of Figure 1 we know that $\Omega(\text{next}(s)) = \Omega(t) = 1$ and that $\mathbb{D}(t) = \text{Mon}$ from the visualization. In this way, the successor segment t can be easily identified and we are able to double-check that both handouts in Figure 1 visualize the rolling stock rotation.

2.2 Handout Quality

The handout function Ω is extensively distributed as a planning tool in the railway industry. In particular, the manual planning of rolling stock rotations is based on similar functions. There, timetabled trips of a planned rotation are equipped with values of $[\mathfrak{v}]$. This is comprehensive because such functions can be made visible (as we have seen for segments) and determine a large part of the rolling stock rotations as well.

Thus, it is not surprising that the function Ω has some expectations in the railway industry. For example, for all segments $s \in S$ arranged in the left of Figure 1

$$\Omega(\text{next}(s)) = (\Omega(s) \mod \mathfrak{v}) + 1 \tag{1}$$

holds. If a successor relation follows equation (1) we call it *logical turn*. It is desired to have as much logical turns as possible in a handout in order to obtain a visualization in that the rolling stock rotations can be easily followed. Imagine that the segments are printed on paper in the order given Ω. In case of a logical turn preceding and succeeding segments are printed next to each other. It is notable that it is not always possible to find a handout such that all successor relations are logical turns. By definition, Ω induces a natural partition of the segments of the rotation into *blocks*. All segments $s \in B$ of a block $B \subseteq \mathcal{S}$ have the same $\Omega(s)$, but a different day of operation. Therefore, each block is of cardinality $|\mathbb{D}|$ and there are always exactly $\mathfrak{v}$ blocks induced by a handout. On the left hand side of Figure 1 the first seven segments $B \subset S$ with $\Omega(s) = 1$ for all $s \in B$ form a block.

Another desired property (if not the most important) is related to the blocks of a handout. For example, on the right hand side of Figure 1 the first seven segments, (i.e., the segments of the first block) are all equal. We say that two segments $u, v \in \mathcal{S}$ with $\Omega(s) = \Omega(t)$ have a *difference* if they cover trips with different train numbers. A handout is desired to have as few differences in its blocks as possible such that patterns can be easily remembered if they are arranged appropriately in blocks. For example, the connection of trips with train number 374 to trips with train number 1061 is reflected in the handout on the right of Figure 1. Note that a handout can only contain fewer differences if appropriate patterns along the rotation exist. This important aspect in rolling stock rotation planning is called *regularity*, see [2] and [3].

2.3 Handout Optimization Problem

Finding Ω is an optimization problem because many handouts of different quality w.r.t. logical turns and differences exist for a given rolling stock rotation. We call this problem *handout optimization problem* (HOP). In order to have a formal reference for this problem, we formulate it as a quadratic assignment problem in this section. To this end, let $x_s^\omega \in \{0,1\}$ be a binary decision variable that takes value one if and only if $\Omega(s) = \omega$ for the segment $s \in S$. In order to qualify handouts we denote by $\operatorname{diff}(s,t) \in \mathbb{Z}_+$ the number of different train numbers in $s \in S$ and $t \in S$ with $s \neq t$. A straight-forward formulation of the HOP as a special quadratic assignment problem ($\mathrm{HOP}_{\mathrm{QAP}}$) reads as follows:

$$\min \sum_{\omega \in [\mathfrak{v}]} \left(\sum_{\substack{s,t \in S \\ s \neq t}} \operatorname{diff}(s,t)\, x_s^\omega x_t^\omega - \alpha \sum_{s \in S} x_s^\omega x_{\mathrm{next}(s)}^{\mathrm{next}(\omega)} \right) \tag{$\mathrm{HOP}_{\mathrm{QAP}}$}$$

$$\sum_{\omega \in [\mathfrak{v}]} x_s^\omega = 1 \qquad \forall\, s \in S, \tag{2}$$

$$\sum_{s \in S(d)} x_s^\omega = 1 \qquad \forall\, d \in \mathbb{D},\ \omega \in [\mathfrak{v}], \tag{3}$$

$$x_s^\omega \in \{0,1\} \quad \forall\, s \in S,\ \omega \in [\mathfrak{v}].$$

Equalities (2) and (3) of program ($\mathrm{HOP}_{\mathrm{QAP}}$) constrain the binary x-variables to perfectly match all segments of S to pairs of $\mathbb{D} \times [\mathfrak{v}]$, i.e., to form a perfect matching (i.e., an assignment) in a bipartite graph that is composed of the two disjoint node parts S and $\mathbb{D} \times [\mathfrak{v}]$. Thus, any feasible solution to program ($\mathrm{HOP}_{\mathrm{QAP}}$) precisely defines a handout Ω where $\Omega(s) = \omega$ if and only if $x_s^\omega = 1$. By the objective function of program ($\mathrm{HOP}_{\mathrm{QAP}}$) we model both the minimization of differences as well as the maximization of logical turns. To this end, quadratic terms are used. Note that these two desired properties compete with each other. A handout that minimizes differences may not maximize the number of logic turns and vice versa, see Figure 1. In order to adjust the relationship between logical turns and differences to a desired level, the parameter $\alpha \in \mathbb{Q}$ is introduced.

▶ **Theorem 1.** *The handout optimization problem is $\mathcal{NP}$-hard even for $\alpha = 0$.*

Proof. This can be proven by a reduction of the the 3-partition problem, which is well known to be $\mathcal{NP}$-hard see [1], to the HOP. ◀

3 Handout Optimization

To solve the handout optimization problem we implemented two solution approaches:

- **Mixed Integer Programming approach:** The model ($\mathrm{HOP}_{\mathrm{QAP}}$) is linearized to a mixed integer programming formulation and solved with a standard MIP solver.
- **Heuristic approach:** We implemented a *construction heuristic*. The procedure is subdivided into two stages. In the first stage, the segments are partitioned into blocks but without assigning a value for Ω to them. In one iteration an assignment problem is set up and solved with an $\mathcal{O}(|V|^3)$ implementation of the classical Hungarian method, see [4]. In the second stage a value of $[\mathfrak{v}]$ is assigned to each of these blocks. Consequently, the segments take over the values assigned to each block in that they are contained. The sum of differences of the segments within the blocks is minimized in the first stage, while the number of logical turns is maximized in the second stage.

4 Computational Results

In this section we present computational results for the optimization of handouts. The considered HOP instances are based on optimized rolling stock rotations provided by DBF.

Table 1 Computational results for handout optimization problems.

$\mathfrak{v}$	columns	rows	δ_1	δ_3	t_1	t_2	t_3	c_1	c_2	c_3
6	4578	1536	0	0	5.61	0.71	4.98	220	221	220
10	21070	4080	1	1	228.52	0.49	447.97	375	377	375
15	70980	8970	11	9	3600.01	0.55	3600.01	368	372	367
18	122598	12816	7	7	3600.02	0.42	3600.06	952	959	951
25	328300	24450	44	9	3600.12	1.25	3600.16	3233	1998	1990
29	512372	32770	∞	7	3600.00	0.49	3600.00	-	1470	1468
31	625828	37386	46	9	3600.08	0.99	3600.07	3371	2006	1991

They differ in characteristics as number of trips, fleet sizes to cover the trips, or possible connections of trips. All computations were performed on CPU with 3.50 GHz, 16 GB of RAM in multi thread mode with four cores using `Gurobi 6.0` with a runtime limit of one hour (3600 seconds). We consider the following three solution approaches to the HOP:

1. static solving of the linearised MIP formulation (HOP_{QAP}) with `Gurobi 6.0`,
2. the construction heuristic,
3. a MIP approach using the solution of approach 2 as warmstart.

Table 1 provides the results of these three solution approaches for our test set. The number of vehicles $\mathfrak{v}$ range from 2 to 31, see the first column of Table 1. As a consequence, we have to solve different models of HOP_{MIP}. The corresponding number of columns and rows of the MIP are given in columns two and three. Columns δ_1 and δ_3 provide the final relative gap for approach 1 and 3, respectively. Note that we explicitly refuse to include δ_2 because without approach 1 it would not be possible to provide a quality measure for approach 2. The last six columns show the computation time t_i and the final cost value c_i for all three approaches $i \in \{1, 2, 3\}$. In all cases the heuristic finds solutions within at most two seconds, see column t_2. The MIP approach allows us to benchmark the heuristic solutions, see the values in columns, c_1 to c_3. This impressively demonstrates the high quality of the solutions provided by approach 2. For the complete set of instances one could observe that approach 3 improves the solution quality after one hour of at most 1%. Thus, we conclude that approach 2 is a very powerful heuristic to solve real world instances of the HOP.

References

1 K. Andreev and H. Räcke. Balanced graph partitioning. *Theory of Computing Systems*, 39(6):929–939, 2006.

2 R. Borndörfer, M. Reuther, T. Schlechte, and S. Weider. A Hypergraph Model for Railway Vehicle Rotation Planning. In Alberto Caprara and Spyros Kontogiannis, editors, *11th ATMOS*, volume 20 of *OASIcs*, pages 146–155, Dagstuhl, Germany, 2011. Schloss Dagstuhl.

3 Boris Grimm, Ralf Borndörfer, Markus Reuther, Thomas Schlechte, and Stanley Schade. Regularity patterns for rolling stock rotation optimization. In *8th International Conference on Applied Operational Research, Proceedings*, volume 8, pages 28 – 32, 2016.

4 H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

[0] The relative gap is defined between the best integer objective UB and the objective of the best lower bound LB as $100 \cdot \frac{UB-LB}{UB}$.

Real-time near-optimal train scheduling and routing in complex railway networks

Marcella Samà[1], Andrea D'Ariano[1], Dario Pacciarelli [1], Francesco Corman[2]

[1] Department of Engineering, Roma Tre University, via della Vasca Navale, 79
– 00146 Rome, Italy. sama@ing.uniroma3.it, dariano@ing.uniroma3.it, pacciarelli@ing.uniroma3.it

[2] Section of Transport Engineering and Logistics, Delft University of Technology, Mekelweg 2
– 2628CD Delft, The Netherlands. f.corman@tudelft.nl

Abstract

This paper focuses on the development of new algorithms for the real-time train scheduling and routing problem in a complex and busy railway network. Since this is a strongly NP-hard problem and practical size instances are complex, simple heuristics are typically adopted in practice to compute feasible but low quality schedules in a short computation time. In order to compute good quality solutions, we consider a mixed-integer linear programming formulation of the problem and solve it with a commercial solver. However, the resolution of this formulation by a commercial solver often takes a too long computation time. Therefore, a new methodology based on the relaxation of some train routing constraints in the formulation is proposed for the quick computation of a good quality lower bound. The lower bound solution is then transformed via a constructive metaheuristic into a feasible schedule, representing a good quality upper bound to the problem. Computational experiments are performed on several disturbed traffic situations for two practical case studies from the Dutch and British railways. The results show that the new lower and upper bounds are computed in a few seconds and are often of similar quality to the ones computed by the commercial solver in hours of computation.

Keywords: Real-Time Railway Traffic Management; Disjunctive Programming; Metaheuristics.

1 Introduction

This work addresses the *Real-Time Train Scheduling and Routing Problem* (RTTSRP), i.e., the problem of computing in real time a conflict-free schedule for a set of trains circulating in a network within a time window $W = [t, t + \delta]$, given the position of the trains at time t and the status of the network in W. The objective function is the minimization of train delays. A schedule is conflict-free if it satisfies the railway traffic regulations, which prescribe a minimum separation between consecutive trains on a shared resource in order to ensure the safety of train movements and to avoid deadlock situations in the network.

The study of real-time train scheduling and routing problems received increasing attention in the literature in the last years. Early approaches tend to solve very simplified problems that ignore the constraints of railway signalling, and that are only applicable for specific traffic situations or network configurations (e.g., a single line or a single junction), see, e.g., the literature reviews in [1, 2, 4, 7, 8, 9, 11]. Among the reasons for this gap between early theoretical works and practical needs are the inherent complexity of the real-time process and the strict time limits for taking and implementing decisions, which leave small margins to a computerized Decision Support System (DSS).

The *alternative graph* of Mascis and Pacciarelli [10] is among the few models in the literature that incorporate, within an optimization framework, the microscopic level of detail that is necessary to ensure the fulfillment of traffic regulations. This model generalizes the job shop scheduling model in order to deal with additional constraints. Each operation denotes the traversal of a resource (block/track section or station platform) by a job (train route).

A big-M Mixed Linear Integer Programming (MILP) formulation of the RTTSRP can be obtained from the alternative graph model by introducing a binary variable for each train ordering decision and a binary variable for each routing decision [12]. The resulting problem is strongly NP-hard [10].

This paper reports on recent improvements implemented in the AGLIBRARY optimization solver [3, 5, 6, 12]. The solver includes a branch and bound algorithm for scheduling trains with fixed routes [5], plus a local search [6], a tabu search [3], and a variable neighborhood search [12] for re-routing trains.

Previous research left open two relevant issues. The first issue is how to certify the quality of the RTTSRP solutions in a short computation time by means of effective lower bounds. This issue is made difficult due to the poor quality of the lower bounds computed by MILP solvers, that are usually based on a linear relaxation of the big-M MILP formulation of the RTTSRP. The second issue concerns with the computation of effective upper bounds through the development of new solution methods. Both issues motivate this paper, whose contribution consists of the following algorithms.

A first algorithm is proposed for the computation of a lower bound for the RTTSRP. This is obtained by the construction of a particular alternative graph for a relaxed RTTSRP in which each train route is composed by two types of components: (i) *real operations* that are in common with all alternative routes of the associated train; (ii) *fictitious operations* that represent the shortest path between two real operations that can be linked by different routing alternatives for the associated train. For the latter type of component, no train ordering decision is modeled, disregarding the potential conflicts between trains. The resulting alternative graph is then solved to optimality by the branch and bound algorithm in [5].

A second algorithm is a constructive metaheuristic proposed in order to optimize the selection of default routes. This metaheuristic starts from the optimal solution obtained for the alternative graph of the relaxed RTTSRP problem and iteratively replaces the fictitious operations of each train with a particular routing alternative. The selection of the routing alternative is based on the evaluation of the insertion of various train routes via the construction of the corresponding alternative graph and the computation of train scheduling solutions via fast heuristics.

Computational experiments are performed on practical-size instances from the Dutch and British railways. The new algorithms often compute good quality lower and upper bounds to the optimal RTTSRP solutions in a shorter computation time compared to a commercial MILP solver.

2 Problem formulation

This section describes our formulation of the RTTSRP. The RTTSRP can be divided into two sub-problems: (i) the selection of a route for each train, and (ii) the train scheduling decisions once the routes have been fixed. We first provide a brief description of the alternative graph model for sub-problem (ii). We then present a big-M MILP formulation for the overall scheduling and routing problem.

2.1 Alternative graph model

The alternative graph (AG) model for sub-problem (ii) of the RTTSRP is a digraph $G = (N, F, A)$ where $N = \{0, 1, ..., n-1, n\}$ is a set of nodes, F is a set of *fixed* arcs, and A a set of pairs of *alternative* arcs.

Each node, except the start 0 and end n nodes, is associated with the start of an operation krj, where k indicates the train, r the route chosen and j the resource it traverses. The start time t_{krj} of operation krj is the entrance time of train k with route r in resource j.

The fixed arcs are used to model running, dwell, connection, arrival, departure, and pass through times of trains. Let the resources p and j be two consecutive resources processed by train k with route r, the fixed arc $(krp, krj) \in F$ models a job constraint between the nodes krp and krj. The weight $w^F_{krp\_krj}$ represents a minimum time constraint between t_{krp} and t_{krj}: $t_{krj} \geq t_{krp} + w^F_{krp\_krj}$. A fixed arc $(umv, krz) \in F$ enforces a connection constraint between train k with route r and train u with route m.

The alternative arcs are used to model the headway times between two consecutive trains. Each pair of alternative arcs $((krj, ump), (umi, krp)) \in A$ models train sequencing decisions between train k with route r and train u with route m on the common resource p. Note that j [respectively i] is the next resource processed by train k [u] when using route r [m]. The two arcs of the pair have associated the weights $w^A_{krj\_ump}$ and $w^A_{umi\_krp}$. In any solution, only one arc of each pair can be selected. If alternative arc (krj, ump) $[(umi, krp)]$ is selected in a solution, the constraint $t_{ump} \geq t_{krj} + w^A_{krj\_ump}$ $[t_{krp} \geq t_{umi} + w^A_{umi\_krp}]$ has to be satisfied. This is to fixing the order of trains, first k and then u [first u and then k].

A solution for sub-problem (ii) of the RTTSRP is represented by the following graph structure. Selection S is a set of alternative arcs obtained by selecting exactly one arc from each alternative pair in A and such that the resulting graph $\mathcal{G}(F,S) = (N, F \cup S)$ does not contain positive weight cycles. The graph selection allows to associate feasible train orders and times to all operations. The objective function is measured as a makespan minimization. Given S and any two nodes krp and uml, we let $l^S(krp, uml)$ be the weight of the longest path from krp to uml in $\mathcal{G}(F,S)$. The start time t_{krp} of $krp \in N$ is the quantity $l^S(0, krp)$, which implies $t_0 = 0$ and $t_n = l^S(0, n)$.

2.2 MILP formulation

A MILP formulation for sub-problem (ii) of the RTTSRP can be obtained from the alternative graph model by representing each fixed arc in F as a linear constraint, by translating each alternative pair in A into a pair of linear constraints, and by introducing a binary variable to model the choice between one of the paired constraints. The variables of sub-problem (ii) are the following: $|N|$ real variables t_{krj} associated to the start time of each operation $krj \in N$, and $|A|$ binary variables $x_{(krj,ump),(umi,krp)}$ associated to each alternative pair $((krj, ump), (umi, krp)) \in A$.

We next extend the MILP formulation for sub-problem (ii) to the overall RTTSRP formulation. The constraint sets F and A are enlarged in order to contain all possible train routing combinations. $|C|$ new binary variables y are associated to the set of routes of the considered train, in addition to the $|N| + |A|$ variables of sub-problem (ii).

The RTTSRP is formulated as the *disjunctive program* (1) with the following notation. M is a sufficiently large number, e.g. the sum of all arc weights. Z is the number of trains, R_b the number of routes for each train $b = 1, ..., Z$. For each train b, only one among the R_b routes can be chosen in any RTTSRP solution. The binary variable y_{ab} indicates if route a is chosen (1) or not (0) for train b. The following constraint holds for train b: $\sum_{a=1}^{R_b} y_{ab} = 1$.

When a route r is chosen for train k (i.e. $y_{kr} = 1$), each fixed constraint in F related to route r and train k must be satisfied. For each fixed arc $(krp, krj) \in F$, $t_{krj} - t_{krp} \geq w^F_{krp\_krj}$ must hold. A fixed arc $(umv, krz) \in F$ enforces a connection constraint between train k with route r and train u with route m (if $y_{um} = y_{kr} = 1$).

$$
\begin{array}{ll}
\min t_n & \\
\sum_{a=1}^{R_b} y_{ab} = 1 & b = 1, ..., Z \\
t_{krj} - t_{krp} \geq w^F_{krp\_krj} + M(1 - y_{kr}) & (krp, krj) \in F \\
t_{krz} - t_{umv} \geq w^F_{umv\_krz} + M(2 - y_{kr} - y_{um}) & (umv, krz) \in F \\
t_{ump} - t_{krj} \geq w^A_{krj\_ump} + M(2 - y_{kr} - y_{um} + x_{(krj,ump),(umi,krp)}) & ((krj, ump), (umi, krp)) \in A \\
t_{krp} - t_{umi} \geq w^A_{umi\_krp} + M(3 - y_{kr} - y_{um} - x_{(krj,ump),(umi,krp)}) & ((krj, ump), (umi, krp)) \in A \\
y_{ab} \in \{0, 1\} & \\
x_{(krj,ump),(umi,krp)} \in \{0, 1\} &
\end{array}
\tag{1}
$$

Regarding the alternative constraints in A, if $y_{um} = y_{kr} = 1$ and the routes m and r of trains u and k use the same resource p of the network, a potential conflict exists on that resource and an ordering decision has to be taken. This is modelled by introducing the binary variable $x_{(krj,ump),(umi,krp)}$ for the alternative pair $((krj, ump), (umi, krp)) \in A$, related to trains u and k travelling on resource p. There are two possible scheduling decisions for each alternative pair $((krj, ump), (umi, krp)) \in A$: if $x_{(krj,ump),(umi,krp)} = 0$ then $t_{ump} - t_{krj} \geq w^A_{krj\_ump}$ must be satisfied (i.e. $(krj, ump) \in S$); if $x_{(krj,ump),(umi,krp)} = 1$ then $t_{krp} - t_{umi} \geq w^A_{umi\_krp}$ must be satisfied (i.e. $(umi, krp) \in S$).

3 The lower bound algorithm

We propose a new method for the computation of a good quality lower bound for the RTTSRP. The basic idea is to reduce the sub-problem (i) to a single route for each train, and solve the resulting sub-problem (ii) to optimality.

Solving the sub-problem (i) requires the construction of the alternative graph $G' = (N', F', A')$ as follows. The set N' contains node 0, node n and the nodes of all jobs. For each train, the corresponding job (train route) starts from node 0 and ends in node n. The intermediate nodes are either real or fictitious operations. We recall that a real operation for a train is an operation in common with all its alternative routes, i.e. an operation to be performed by the train in any solution of sub-problem (i). A fictitious operation is the shortest path between two real operations that can be linked by different routing alternatives.

F' is the set of fixed arcs in G'. In general, the weight $w^{F'}_{krp\_krj}$ of each arc $(krp, krj) \in F'$ corresponds to the less stringent constraint between nodes krp and krj when looking at all routing alternatives of k. In particular, if krp is a fictitious operation and krj is a real operation, the weight $w^{F'}_{krp\_krj}$ is equal to the minimum travel time between the start of krp and the start of krj, taken from the shortest path among the routing alternatives of k. The weight $w^{F'}_{umv\_krz}$ corresponds to the less stringent connection constraint between the nodes umv and krz when looking at all routing alternatives of u and k.

A' is the set of alternative pairs in G'. In each pair $((krj, ump), (umi, krp)) \in A'$, ump and krp are necessarily real operations, since the lower bound method only takes scheduling decisions between conflicting resources in all train routing combinations. In other words, if a potential conflict can be avoided by some routing combinations, this conflict is disregarded in the lower bound method.

Sub-problem (ii) requires the computation of the optimal graph selection S^* for G'. This corresponds to solving the following *disjunctive program* (2). The lower bound value is the optimal solution of sub-problem (ii). In this paper, we solve sub-problem (ii) via the branch and bound algorithm in [5], truncated at a given time limit of computation. In case the optimal solution cannot be found, the algorithm always returns a valid lower bound value.

$$
\begin{aligned}
&\min t_n && \\
&t_{krj} - t_{krp} \geq w^{F'}_{krp\_krj} && (krp, krj) \in F' \\
&t_{krz} - t_{umv} \geq w^{F'}_{umv\_krz} && (umv, krz) \in F' \\
&t_{ump} - t_{krj} \geq w^{A'}_{krj\_ump} + Mx_{(krj,ump),(umi,krp)} && ((krj, ump), (umi, krp)) \in A' \\
&(t_{krp} - t_{umi} \geq w^{A'}_{umi\_krp} + M(1 - x_{(krj,ump),(umi,krp)}) && ((krj, ump), (umi, krp)) \in A' \\
&x_{(krj,ump),(umi,krp)} \in \{0, 1\} &&
\end{aligned}
\tag{2}
$$

4 The upper bound algorithm

We propose a constructive metaheuristic for the computation of a good quality upper bound for the RTTSRP. This metaheuristic starts from the alternative graph G' and a selection S generated by the lower bound method, and iteratively trasforms each job (train route) with some fictitious operations into a new job with real operations only. At each iteration, the train to be trasformed is chosen from a train list, according to a sort criteria defined in advance (in this paper we order the trains on a first come first served basis). The selection of a real route for the train is based on a local search procedure in which we evalute the insertion in G' of all its potential real routes by an alternative graph greedy heuristic. After the local search, the real route generating less consecutive delays is inserted in G', and the chosen alternative graph heuristic updates the selection S of G'. When all trains with some fictitious operations have been processed, the train list is empty and the iterative process ends. The metaheuristic returns $l^S(0, n)$ obtained in G' at the last iteration, that corresponds to a solution to the RTTSRP.

5 Computational experiments

This section presents the experimental results on the two practical case studies: a British test case, part of the East Coast Main Line of The United Kingdom; and a Dutch test case, the Utrecht - Den Bosch railway network. Both the test cases are modelled with a microscopic level of detail, which means that switches, signals, block sections, and track segments in complex station areas are included (yielding several hundreds of resources). Also, train movements are described with a precision of seconds.

For each case study, we consider a set of 15 RTTSRP instances, varying the initial delays of trains. The experiments are executed on a workstation Power Mac with processor Intel Xeon E5 quad-core (3.7 GHz), 12 GB of RAM. We compare the lower and upper bounds obtained of Sections 3 and 4 implemented in AGLIBRARY, and the MILP formulation of Section 2.2 solved by IBM LOG CPLEX MIP 12.0. In the lower bound method, we use the Branch and Bound (BB) algorithm of [5] truncated at 1 hour of computation. However, for all the tested instances the time limit of BB is never reached. The CPLEX solver is truncated at 2 hours of computation. The CPLEX time limit is reached for 16/30 instances.

Table 1 presents average information on the 15 RTTSRP instances of each test case. This network presents a huge number of $|A|$ variables, since these are defined for each pair of trains sharing a resource in one hour traffic prediction horizon and for all train routing combinations. For the British test case we only consider two routes per train, and thus less $|A|$ and $|C|$ variables compared to the Dutch case study.

Table 1: Characteristics of the instances

Case Study	Network Length (km)	Num Trains	Num Routes Per Train	Num Resources Per Train	MILP Variables $\|N\|$	$\|A\|$	$\|C\|$
Dutch	50	40	9	31	1615	1092557	356
British	80	90	2	69	5565	46219	128

Table 2 reports the average results for the 15 RTTSRP instances of each test case, in terms of Lower Bound (LB) and Upper Bound (UB) values (in seconds) obtained by the two tested solvers, and the time (in seconds) required to compute those values.

Table 2: Average computational results

Case Study	Solver	LB Value	LB Comp. Time	UB Value	UB Comp. Time
Dutch	AGLIBRARY	94.1	0.1	179.4	11.5
Dutch	CPLEX	58.7	244.9	1259.9	5399.6
British	AGLIBRARY	1023.9	4.0	1472.1	7.8
British	CPLEX	1013.3	469.9	1062.8	872.5

The results on the Dutch case study show that AGLIBRARY, on average, outperforms CPLEX both in terms of LB quality, UB quality and the related computation times.

Regarding the results of the British case study, the LB quality of AGLIBRARY is, on average, superior to the one of CPLEX, while the constructive metaheuristic is not always able to compute an UB value close to the one obtained by the other solver. However, AGLIBRARY always requires a few seconds to compute the LB and UB values, while CPLEX requires several minutes of computation and is thus not applicable to recover real-time traffic disturbances.

6 Conclusions and future research

This paper introduces new methods for the computation of good quality LB and UB for the RTTSRP. The LB value is computed by relaxing some constraints in the MILP formulation of the RTTSRP and solving the relaxed formulation via the BB algorithm of [5]. The UB value is obtained by transforming the LB solution in a RTTSRP solution via greedy heuristics. The computational results are promising, since both the new LB and UB values are computed very quickly and are often of good quality compared to the LB and UB values computed with CPLEX in hours of computation.

Further research should be dedicated to incorporating the new LB and UB methods in advanced heuristic, metaheuristic and exact algorithms for the RTTSRP.

References

[1] Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J. (2014) An overview of recovery models and algorithms for real-time railway rescheduling. *Transp. Res. Part B* **63**, 15–37.

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

[3] Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M. (2010) A tabu search algorithm for rerouting trains during rail operations. *Transp. Res. Part B* **44** (1), 175–192.

[4] Corman, F., Meng, L. (2014) A Review of Online Dynamic Models and Algorithms for Railway Traffic Control. *IEEE Trans. on Intelligent Transportation Systems* **16** (3), 1274–1284.

[5] D'Ariano, A., Pacciarelli, D., Pranzo, M. (2007) A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research* **183** (2), 643–657.

[6] D'Ariano, A., Corman, F., Pacciarelli, D., Pranzo, M. (2008) Reordering and Local Rerouting Strategies to Manage Train Traffic in Real Time. *Transportation Science*, **42** (4), 405–419.

[7] Fang, W., Yang, S., Yao, X. (2015) A Survey on Problem Models and Solution Approaches to Rescheduling in Railway Networks. *IEEE Trans. on Intelligent Transport. Systems* **16** (6), 2997–3016.

[8] Hansen, I.A., Pachl, J. (2014) Railway Timetabling & Operations. Eurailpress, Hamburg, Germany.

[9] Lusby, R.M., Larsen, J., Ehrgott, M., Ryan, D.M. (2013). A set packing inspired method for real-time junction train routing. *Computers and Operations Research* **40** (3), 713–724.

[10] Mascis, A., Pacciarelli, D. (2002) Job shop scheduling with blocking and no-wait constraints. *European Journal of Operational Research* **143** (3), 498–517.

[11] Pellegrini, P., Rodriguez, J. (2013) Single European sky and single European railway area: A system level analysis of air and rail transportation. *Transp. Res. Part A* **57** (1), 64–86.

[12] Samà, M., D'Ariano, A., Corman, F., Pacciarelli, D. (2017) A variable neighborhood search for fast train scheduling and routing during disturbed railway traffic situations. *Computers and Operations Research* **78**, 480–499.

Multimodal Transportation Services
SB1: Multiobjective and Multidimensional Logistics
Saturday 11:15 – 12:45 PM
Session Chair: Olivier Péton

11:15 Multi-Criteria Decision Making when Planning & Designing Sustainable Multi-Modal Transportation in a Corridor
Marie Louis, Eric Gonzales*
University of Massachusetts Amherst

11:45 Bin-Packing Problems with Load Balancing and Stability Constraints
Alessio Trivella, David Pisinger*
Technical University of Denmark

12:15 Multi-Directional Local Search for A Bi-Objective Vehicle Routing Problem with Lexicographic Minimax Load Balancing
[1]Fabien Lehuédé, [1]Olivier Péton, [2]Fabien Tricoire*
[1]IMTA-LS2N Nantes, [2]University of Vienna

Multi-Criteria Decision Making when Planning & Designing Sustainable Multi-modal Transportation in a Corridor

Marie P. Louis and Eric J. Gonzales

Public transit is often promoted as a way to reduce congestion and transportation-related emissions in cities. To design transit service for a multimodal corridor in which people may also driver cars, there are three objectives to consider: minimize generalized cost to users (including direct costs to users and travel time), minimize agency cost for infrastructure and operations, and minimize emissions of pollutants such as greenhouse gas emissions. Analytical models of these three types of costs have been developed to compare private cars, buses operating in mixed traffic, buses in dedicated lanes, light rail in mixed traffic (i.e., tram), light rail in dedicated right-of-way, and grade-separated metro. A Pareto analysis is conducted to reveal the trade-off between cost and emissions when optimizing the stop spacing and service headway for the deployment of each mode to serve demand characterized by density of trip-generation, average trip length, and average value of time. The models have a general structure that allows for comparison across many transit modes, and the results allow for systematic evaluation of the sustainability of multimodal corridors under different demand conditions. This study plugs a research gap by directly targeting the effect that greenhouse gas emissions have on the design of transit service. Although emission costs do not have a big effect on the optimal design of transit service for a specific mode, it can have important consequences for mode selection and planning incentives for travelers to use transit. These models also provide estimates of total emissions in the corridor.

Bin-packing problems with load balancing and stability constraints

Alessio Trivella*, David Pisinger
Department of Management Engineering
Technical University of Denmark, Kgs. Lyngby, Denmark
*Email: atri@dtu.dk

1 Introduction

The *Bin-Packing Problem* (BPP) is one of the most investigated and applicable combinatorial optimization problems. The problem consists of packing objects of different sizes into a finite number of similar bins, such that the number of used bins is minimized. Applications of the bin-packing problem apper in a wide range of disciplines, including transportation and logistics, computer science, engineering, economics and manufacturing. The problem is well-known to be $\mathcal{NP}$-*hard* and difficult to solve in practice, especially when dealing with the multi-dimensional cases.

Closely connected to the BPP is the *Container Loading Problem* (CLP), which addresses the optimization of a spacial arrangement of cargo inside a container or transportation vehicle, with the objective to maximize the value of the cargo loaded or the volume utilization. The CLP focuses on a single container, and has been extended in the literature to handle a variety of different constraints arising from real-world problems. Consider for example the problem of arranging items into an aircraft cargo area such that the barycenter of the loaded plane is as close as possible to an ideal point given by the aircraft's specifications. The position of the barycenter has an impact on the flight performance in terms of safety and efficiency, and even a minor displacement from the ideal barycenter can lead to a high increase of fuel consumption [1]. Similar considerations apply when loading trucks and container ships.

The aim of this work is to integrate realistic constraints related to e.g. load balancing, cargo stability and weight limits, in the multi-dimensional BPP. The BPP poses additional challenges compared to the CLP due to the supplementary objective of minimizing the number of bins. In particular, in section 2 we discuss how to integrate bin-packing and load balancing of items. The problem has only been considered in the literature in simplified versions, e.g. balancing a single bin or introducing a feasible region for the barycenter. In section 3 we generalize the problem to handle cargo stability and weight constraints.

2 The integrated packing and balancing problem

Packing and balancing the load of a set of items represent two conflicting objectives and it can be convenient to incorporate them in a single problem. In [2] we develop an integrated approach for solving the multi-dimensional bin-packing problem with load balancing. The goal is to arrange the items in the smallest number of bins, while ensuring the best overall load balancing of the used bins, i.e. ensuring that the average barycenter of the loaded bins falls as close as possible to an ideal point, for instance the center of the bin. We use:

- a Mixed-Integer Program (MIP) model. Despite having a quadratic number of variables and constraints, the model is considerably more complex than the BPP as several new sets of support binary variables and conditional constraints are necessary to optimize load balancing. The model is able to solve only instances involving up to 20 items.
- a multi-level local search heuristic able to deal with large instances. The algorithm takes advantage of the Fekete-Schepers representation of feasible packings in terms of particular classes of interval graphs [3], and iteratively improves the load balancing of a bin-packing solution using three different search levels:
    1. The first level explores the space of transitive orientations (TROs) of the complement graphs associated with the packing. This is made possible by exploiting nice theoretical properties of interval graphs and a relation TRO-packing.
    2. The second level modifies the structure of the interval graphs.
    3. The third level exchanges items between bins by repacking proper n-tuples of weakly balanced bins, coded inside a variable neighborhood search framework.

Extended computational experiments can be found in [2]. The results reveal that an effective load balancing can be obtained, with a running time rarely exceeding 3-5 minutes even for difficult instances with up to 200 items (C program on an Intel Core i5 with 8GB RAM). When the optimal barycenter is the geometric center of the bin, for instance, more than 95% of the initial imbalance can be removed by using the three search phases.
In Table 1 is a summary of the main features of the two approaches and the differences between them. It will be useful when considering additional constraints in the next section.

Table 1: MIP vs. local search.

	MIP	Local search
algorithm type:	exact	heuristic
objective function:	only L_1	any, e.g. L_p
ideal barycenter:	any	any
instances handled:	small	medium/large

3 Including stability and weight constraints

In practical contexts, additional properties are relevant when packing containers to ensure the stability of the cargo (see e.g. [4, 5]). Mechanical stability can be enforced at a *static* or *dynamic* level, related to obtaining a stable motionless or moving cargo, respectively. Moreover, a number of weight constraints may apply depending on the specific transport application. In the following, we limit the discussion to static stability, which is by itself challenging, and provide an overview of the different stability and weight requirements considered.

Table 2: Static stability and weight constraints.

Property	Description
Lower gapless	Each box touches another box (or the bin) below it
Full base support	The entire base of each box is in contact to other boxes (or the bin)
Partial base support	A fix percentage of the box base is in contact to other boxes (or bin)
Barycenter support	The box barycenter is located above the surface of a supporting box
Mechanical equilibrium	The sum of external forces and torque acting on each box is zero
Load bearing	Maximum pressure that can be applied over the top face of a box
Weight limit	Items packed in each bin cannot exceed a given maximum weight
Weight distribution	The weight is distributed within the bin according to certain criteria

In the previous section, we provided two tools to solve different instances of the load-balanced bin-packing problem: an MIP model and a heuristic based on local search. The goal now is to understand whether it is possible (and straightforward) to embed the additional constraints of Table 2 into the two algorithms. In the remainder of the section, we briefly discuss two of the properties; the others will be examined more in detail in the paper following this abstract.

3.1 Gapless property

The local-search algorithm can be easily adapted to lower gapless requirements. Indeed, each TRO of the z-interval graph complement can be associated to a lower gapless packing. In other words, we can simply collapse the bunch of packings associated to a TRO to a single lower gapless solution, narrowing in this way the search space.

The gapless property is, in contrast, very hard to model in the MIP framework. However, if the ideal barycenter location is the center of the bin's base, than enforcing the lower gapless property is actually not needed, as an optimal solution necessarily satisfies it. If not, it would be possible to move downwards at least one box in one bin, keeping the others fixed, decreasing the total cost.

3.2 Weight limit

The MIP model contains binary variables $c_{ij} = 1$ if and only if item i is in bin j. Using these variables, it is easy to identify the total load of a bin and impose a limit by adding for each bin j the constraints $\sum_i c_{ij} m_i \leq \mathcal{W}^{max}$, where m_i is the mass of item i. The weight limit can also be formulated here as bin-dependent $\mathcal{W}^{max} = \mathcal{W}_j^{max}$.
Integrating weight limits in the heuristic algorithm affects both the construction phase and external search (recombination of bins). The two phases are in a sense similar, as the external search consists of applying a constructive heuristic to a certain number of bins. Thus, both phases can be modified by closing earlier bins where no more boxes can be added without violating the weight limit. If n bins cannot be recombined in the same number n of bins, each fulfilling the weight limit, we discard the solution and go to the next local search move.

3.3 Further work

Future work includes quantifying the impact of additional stability constraints on the volume utilization of containers and on the objective function of the integrated load-balanced bin-packing problem.

References

[1] Mongeau M. and Bes C., Optimization of aircraft container loading, *IEEE Transactions on Aerospace and Electronic Systems*, 39, 140-150, 2003.

[2] Trivella A. and Pisinger D., The load-balanced multi-dimensional bin-packing problem, *Computers & Operations Research*, 74, 152-164, 2016.

[3] Fekete S.P. and Schepers J., A Combinatorial Characterization of Higher-Dimensional Orthogonal Packing, *Mathematics of Operations Research* 29(2), 353-368, 2004.

[4] Ramos A.G, Oliveira J. F., Gonçalves J.F. and Lopes M.P., A container loading algorithm with static mechanical equilibrium stability constraints. *Transportation Research Part B*, 91, 565-581, 2016.

[5] Junqueira L., Morabito R. and Yamashita D.S., Three-dimensional container loading models with cargo stability and load bearing constraints, *Computers & Operations Research*, 39, 74-85, 2012.

Multi-directional local search for a bi-objective vehicle routing problem with lexicographic minimax load balancing

Fabien Lehuédé[1], Olivier Péton[1], Fabien Tricoire[2]

1. Institut Mines-Telecom Atlantique,

Laboratoire des Sciences du Numérique de Nantes, UMR CNRS 6004, Nantes, France

2. Department of Business Administration, University of Vienna, Vienna, Austria

1 Introduction

Vehicle routing problems are mostly solved to design minimum cost routes. In some cases, generally when designing the routes for a given private fleet, a company may be ready to sacrifice savings on cost to better balance the workload over its drivers. In such cases, the load of a driver or a vehicle is generally measured by the length, cost or duration of a route. Accordingly, the length of the longest route has been taken as equity measure and some min-max approaches have been proposed [Golden et al., 1997]. An unfortunate consequence of this approach is that many solutions can have the same equity measure once the length of the longest route has been found. Hence, significantly different solutions are considered equivalent and other equity measures have been proposed. One of the most studied measure in VRP, called *range* in Matl et al. [2016], considers the difference between the longest and the shortest routes in the solution [Jozefowiez et al., 2002]. Nevertheless this measure is not monotonic: it can be improved by increasing, even in an inconsistent way, the length of the shortest route of a solution without decreasing the length of the others. In a recent computational study where several route balancing models are compared, Halvorsen-Weare and Savelsbergh [2016] provide many examples of so-called *artificially balanced* solutions that are considered to be good when using the range objective. Matl et al. [2016] present a very interesting study of equity models in VRPs, listing the desirable properties that should be satisfied by so-called *inequality measures*. They compare a broad set of measures, both looking at their mathematical properties and based on a numerical analysis on small CVRP instances. This analysis includes the lexicographic minimax approach and shows its good properties.

As recently summarized by Ogryczak et al. [2014] in a survey on *fair optimization and networks*, the lexicographic minimax approach has been used in domains such as network optimization, facility location and network optimization to produce equitable or fair solutions. This approach was introduced under the name leximin (in a welfare maximization context) in social choice theory by Sen [1970]. It has also been widely used under this name in constraint programming, to model equity in combinatorial optimization problems [Bouveret and Lemaître, 2009].

The lexicographic minimax refines the min-max approach: informally speaking, when a minimal value has been found for the longest route, the lexicographic minimax considers the second longest route, the third longest route, and so on, until all ties have been broken. Mathematically, let us denote $(t_1, ..., t_m)$ the vector of route lengths of a solution. Let σ denote a permutation of indices such that routes are ordered in decreasing order of length (ie. such that $t_{\sigma(1)} \geq ... \geq t_{\sigma(m)}$). Then a solution with route vector $(t_1, ..., t_m)$ is a lexicographic minimax solution if for any other solution with route lengths $(t'_1, ..., t'_m)$:

$$(t_{\sigma(1)}, ..., t_{\sigma(m)}) \leq_{lex} (t'_{\sigma(1)}, ..., t'_{\sigma(m)}).$$

In this paper, we consider the problem of solving the CVRP with two objective functions: the sum of routing costs and the lexicographic minimax over routes costs. One complexity for optimization algorithms is that the lexicographic minimax approach does not establish an evaluation of how balanced a solution is. It is based on a binary relation which allows to compare two solutions and determine if one is more balanced than the other. However, we believe that this binary relation defines a dominance relation that can easily be integrated in VRP algorithms.

2 Solution method

We propose to integrate the lexicographic minimax approach in a multi-objective optimization framework called Multi-Directional Local Search (MDLS) [Tricoire, 2012].

2.1 MDLS principle

MDLS offers a very simple local search framework but it still competes with state-of-the-art methods when solving multi-objective optimization problems. In MDLS, a local search LS_j is defined for each objective j. This local search is later performed in order to improve solutions with respect to objective j. A set of non-dominated solutions is kept in an archive and returned at the end of the algorithm. An iteration consists in (i) selecting a solution from the archive, (ii) performing local search on this solution for each objective/direction, thus producing a new feasible solution in each direction and (iii) updating the archive using newly produced solutions.

2.2 Local search components

In our algorithm, we consider that local search consists of one Large Neighborhood Search (LNS) iteration. Several ruin and recreate operators are defined for each objective. Hence, at each iteration, for each objective (i) a ruin and a recreate operator are randomly selected in the set of operators for that objective and (ii) a new solution is produced using the selected operators. The *ruin quantity*, used in the destroy operator, may be operator-dependent.

The *cost operators* are defined according to the classical LNS operators for the VRP [Pisinger and Ropke, 2007]. The set of ruin operators that we use are: *random removal, worst removal, related removal* and *route removal*. The recreate operators for the cost objective are the *cheapest insertion heuristic* and the *k-regret heuristic* for $k = 2, 3, 4$.

The main contribution of this work is to introduce *lexicographic minimax operators*. They constitute rather simple extensions of the relevant classical operators to the lexicographical minimax approach. The ruin operators include the *random removal* and the *related removal* as well as the following two operators:

- *worst max removal:* at each iteration this operator removes, from the longest route, the customer that decreases the most the length of this route, until the number of removed customers is equal to the ruin quantity.

- *longest route removal:* in this operator, all customers from the longest route are removed, until the number of removed customers is greater than or equal to the ruin quantity. Several routes may be destroyed this way.

Two sets of recreate operators have been designed to guide the search towards lexicographic minimax efficient solutions:

- The *lexicographic minimax cheapest insertion* and *lexicographic minimax k-regret* extend the classical cheapest insertion heuristic and the k-regret heuristic according to the lexicographic minimax approach.

- The *min-max cheapest insertion* and *min-max k-regret* extend these heuristics, but using only the length of the longest route to guide the search and the solution cost increase to break ties.

Besides proposing these extensions, as the lexicographic minimax heuristics involve sorting route vectors to compare solutions, we want to assess whether the faster min-max heuristics can be more efficient to guide the search towards lexicographic minimax solutions.

2.3 Experiments

The designed MDLS is evaluated on the Christofides CVRP instances, which are traditionally used to benchmark the VRP with load balancing.

We test three configurations, which all include cost operators and lexicographic minimax ruin operators. Configuration *leximax* integrates the *lexicographic minimax cheapest insertion* and *lexicographic minimax k-regret*. Configuration *max* integrates *min-max cheapest insertion* and *min-max k-regret*. Configuration *all* includes all of these recreate operators for the lexicographic minimax objective.

MDLS being stochastic, 10 runs are performed for each configuration and each instance. We execute 1-minute as well as 30-minute runs.

For each instance, a reference set is constructed by taking the non-dominated union of the sets returned by each run for each configuration (3×10 fronts in total). We first look at the performance of each configuration on 1-minute runs. We consider two indicators: the percentage of solutions from the reference front found in a given run, and the percentage of solutions found which are within 1% of a solution from the reference front. A solution x_1 is within 1% of another solution x_2 if, when the cost and all route lengths of x_2 are multiplied by 1.01, then x_1 dominates this transformed solution. This information is summarized in the two plots from Figure 1. On the left plot, for each instance and each configuration, we represent the percentage of solutions of the reference front that are found on each run. The distribution of the performance evaluation of each run is displayed with a Box Plot. Apart from the vrpnc1 instance, it is clear that the percentage of solutions found for each run remains quite low. Comparing the three configurations, it is not possible to clearly state that one dominates the other on these experiments. The right plot shows the percentage of solutions of the reference front that lie within a 1% distance of a solution returned by each run. This plot shows that on a short runtime, although a large proportion of the reference front is not found on each run, the returned approximation remains within close distance of this reference front for most instances and most configurations. Again, we cannot state that a configuration dominates the other.

Figure 2 presents the same Box Plots for a set of 30-minute runs. On the left plot, the proportion of solutions of the reference front found on each separate run is slightly improved, but remains

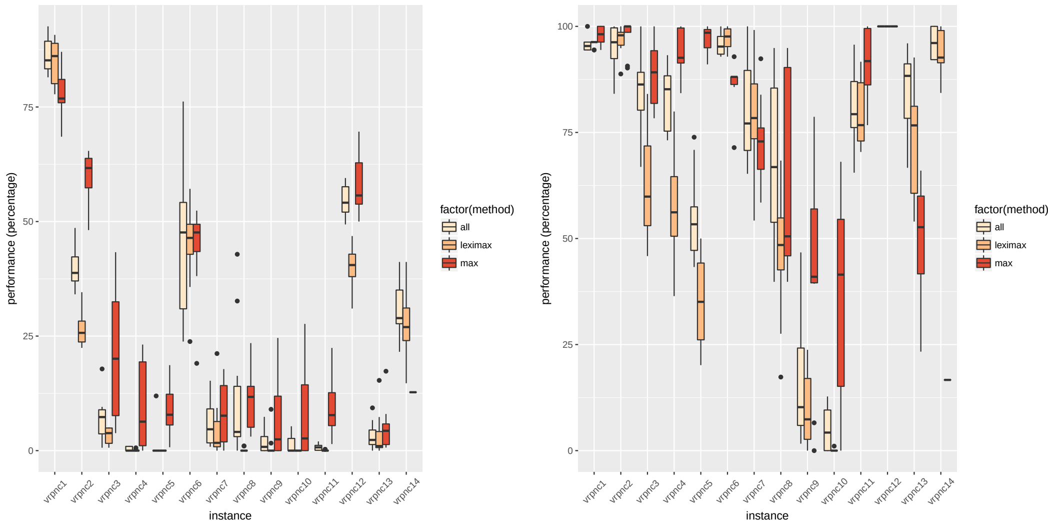


Figure 1: Boxplots charts for the number of solution of the reference front that have been found and within 1% distance for each run and each instance in one minute.

quite low. When looking at the percentage of solutions within a 1 % distance of the returned approximation, the performance is clearly good for all instances but vrpnc9 and vrpnc10. Although configuration *all* gives a better performance on average over all instances, it is dominated by configuration *max* on the vrpnc10 instance. It can also be pointed out that, except for instances vrpnc13 and vrpnc14, the simple *max* configuration remains competitive with the more elaborate *leximax* and *all* configurations.

3 Conclusion

We performed an adaptation of classical insertion heuristics for the lexicographic minimax; preliminary experiments show that the search for efficient solutions could be guided by insertion heuristics based on a min-max criterion.

We believe that the lexicographic minimax approach is an interesting alternative that should be considered to model load balancing in vehicle routing problems.

References

S. Bouveret and M. Lemaître. Computing leximin-optimal solutions in constraint networks. *Artificial Intelligence*, 173(2):343–364, 2009.

B. L. Golden, G. Laporte, and É. D. Taillard. An adaptive memory heuristic for a class of vehicle routing problems with minmax objective. *Computers & Operations Research*, 24(5):445–452, 1997.

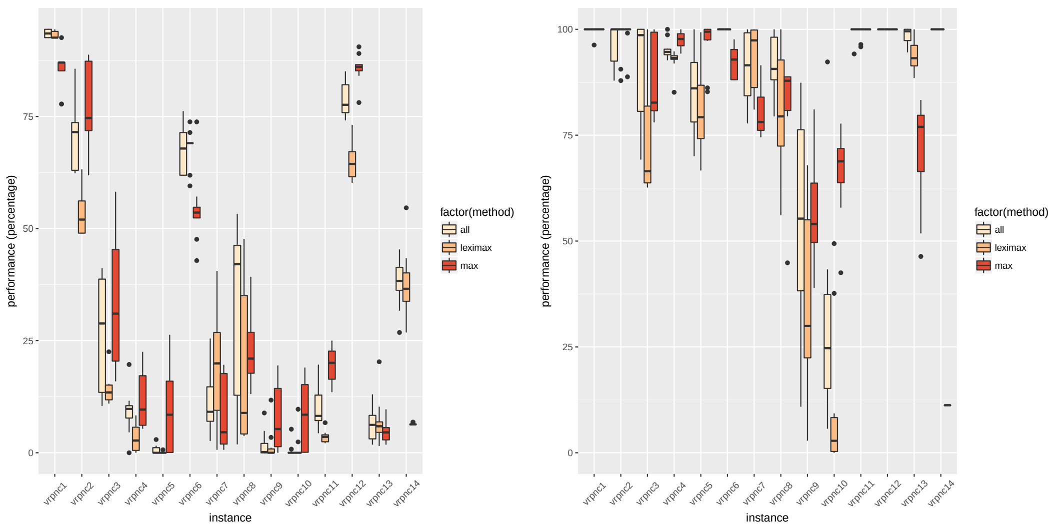


Figure 2: Boxplots charts for the number of solution of the reference front that have been found and within 1% distance for each run and each instance in 30 minutes.

E. E. Halvorsen-Weare and M. W. Savelsbergh. The bi-objective mixed capacitated general routing problem with different route balance criteria. *European Journal of Operational Research*, 251(2): 451–465, 2016.

N. Jozefowiez, F. Semet, and E.-G. Talbi. Parallel and hybrid models for multi-objective optimization: Application to the vehicle routing problem. In *International Conference on Parallel Problem Solving from Nature*, pages 271–280. Springer Berlin Heidelberg, 2002.

P. Matl, R. F. Hartl, and T. Vidal. Equity objectives in vehicle routing: A survey and analysis. *arXiv preprint arXiv:1605.08565*, 2016.

W. Ogryczak, H. Luss, M. Pióro, D. Nace, and A. Tomaszewski. Fair optimization and networks: A survey. *Journal of Applied Mathematics*, 2014, 2014.

D. Pisinger and S. Ropke. A general heuristic for vehicle routing problems. *Computers & operations research*, 34(8):2403–2435, 2007.

A. K. Sen. *Collective choice and social welfare.* North Holland, 1970.

F. Tricoire. Multi-directional local search. *Computers & operations research*, 39(12):3089–3101, 2012.